VOL. XV FEBRUARY, 1931 No. 1

The Journal of
APPLIED PSYCHOLOGY

JAMES P. PORTER, EDITOR
Ohio University, Athens, Ohio
WITH A BOARD OF CO-OPERATING EDITORS

TABLE OF CONTENTS

MAXINE BAUMGARTNER. The Correlation of Direct Suggestibility with Certain Character Traits 1
HENRY F. ADAMS. The Theory of Two Factors: An Alternative Explanation. Part I: General Factor, "G" 16
M. EUSTACE BROOM, JOSEPHINE DOUGLAS, AND MARION RUDD. On the Validity of Silent Reading Tests 35
VERNA KIMBERLEY WALLACE AND JOHN L. BROSE. Who Knows About Aviation? 39
C. M. LOUTTIT. Test Performance of a Selected Group of Part-Hawaiians 43
ROSALIND L. STREEP. A Comparison of White and Negro Children in Rhythm and Consonance 53
MILES A. TINKER AND DONALD G. PATERSON. Studies of Typographical Factors Influencing Speed of Reading. V. Simultaneous Variation of Type Size and Line Length 72
EDWARD A. LINCOLN. A Reliability of the Lincoln Hollow Square Form Board, and a Comparison of Hollow Square Scores with Stanford Binet Mental Ages 79
HERBERT W. ROGERS. Motion Study Tables for Clerical Operators 82
MARION HAMMOND. A Note Concerning Rhythm Tests in Poetry and in Music 90
B. F. HAUGHT. The Language Difficulty of Spanish-American Children 92
C. H. ESTABROOKS. Personality Training and the Changing Emphasis in the Liberal Arts College 96
NOTES AND NEWS 100
BOOK REVIEWS 103
NEW BOOKS AND PAMPHLETS 107

PUBLISHED BI-MONTHLY BY THE EDITOR
All communications should be addressed to the Editor,
Ohio University, Athens, Ohio

Printed by
WAVERLY PRESS, INC.
Baltimore, Md.

Entered as second class matter June 3, 1924, at the post office at Baltimore, Maryland, under the act of March 3, 1879.
Acceptance for mailing at special rate of postage provided for in section 1103, act of October 3, 1917. Authorized December 21, 1922.

COPYRIGHT 1931, BY JAMES P. PORTER

The Journal of Applied Psychology

CO-OPERATING EDITORS

W. F. Book, Indiana University; W. V. Bingham, Personnel Research Federation, New York; June E. Downey, University of Wyoming; Robert H. Gault, Northwestern University; William Healy, Baker Foundation, Juvenile Court, Boston; H. L. Hollingworth, Columbia University; Joseph Jastrow, University of Wisconsin; Herbert S. Langfeld, Princeton University; James H. Leuba, Bryn Mawr College; Henry C. McComas, Princeton University; Walter D. Scott, Northwestern University; C. E. Seashore, University of Iowa; Daniel Starch, Boston, Mass.; Edward K. Strong, Jr., Leland Stanford Jr. University; L. M. Terman, Leland Stanford Jr. University; E. L. Thorndike, Columbia University; J. R. Wallace Wallin, Department of Education, Baltimore, Md.; G. M. Whipple, Denver, Mass.; Helen Thompson Woolley, Teachers College, Columbia University.

The scope of The Journal of Applied Psychology will include the following: (a) The application of psychology to vocational activities, such as law, art, public speaking, industrial and commercial work, and problems of business appeals. (b) Studies of individual mentalities, such as types of character, special talents, genius, and individual differences, including the problems of mental diagnosis and vocational prognosis. (c) The influence of general environment conditions, such as climate, weather, humidity, temperature; also such conditions as nutrition, fatigue, etc. (d) The psychology of everyday activities, such as reading, writing, speaking, singing, playing games or playing musical instruments, sports, etc.

Communications, manuscripts, reviews, summaries, notices, books and off-prints for review should be addressed to Dr. James P. Porter, Ohio University, Athens, Ohio. Contributors are requested to follow the conventional method of citation employed by this Journal. In every case the year, volume and pages should be specified.

The Journal of Applied Psychology is published bi-monthly. The subscription price is $4.00 a year. Single numbers $1.00. Remittances and business communications should be addressed to the Editor, James P. Porter, Ohio University, Athens, Ohio.

THE CORRELATION OF DIRECT SUGGESTIBILITY WITH CERTAIN CHARACTER TRAITS[1]

MAXINE BAUMGARTNER

University of Wisconsin

THE PROBLEM

The present investigation is a part of a program of experimentation concerned with the psychology of suggestion and suggestibility. It has seemed likely on the basis of general observation that direct suggestibility might not necessarily be an undesirable trait, as is so widely supposed. On the contrary it was regarded as possible that this form of suggestibility might even be an important component of a number of thoroughly desirable traits of character. Among these may be mentioned optimism, sympathy, tact, sincerity, and amiability. The investigation was designed to test this hypothesis experimentally. In addition there was a particular interest in determining whether the sample of behavior measured by the test of direct suggestibility here employed really measures the trait in everyday life called suggestibility on the one hand and negativism on the other.

[1] From the Psychology Laboratory, University of Wisconsin. The author wishes to express her thanks to the five supervisors who gave so generously of their time in judging the character traits and to the seventy pupil nurses who patiently went through the test. The measurements of the smoked-paper records were performed by Mr. Wilbert S. Ray. The laborious computations of the correlation coefficients were performed at the Institute of Human Relations, Yale University by Dr. G. W. Williams and Mr. Robert G. Krueger. The expense of measuring the records was borne by a fund from the Social Science Research Council. Prof. Clark L. Hull has directed the work and has aided in the interpretation of results.

The method, in brief, was to determine the correlation existing between the direct suggestibility test scores of a large group of subjects with the composite scores obtained from rather elaborate rankings in respect to the several character traits. No illusions were entertained either as to the unitary (i.e., generalized) nature of such character traits or as to the accuracy of obtainable rankings of such composites. The work was undertaken in the hope, despite the intangible nature of such variables and the attendant difficulties of measurement, that the dim outlines of a relationship might appear if such exist. Inaccuracies and distortions in measurement attenuate observed relationships below their true strength though they do not ordinarily make them entirely disappear.

SUBJECTS

The subjects employed in the investigation were 56 pupil nurses in the training school run in connection with Memorial Hospital which is a part of the University of Wisconsin Medical School. Seventy subjects were tested, but the records of 14 had to be discarded because of some inadequacy of the data, mainly because they were not ranked by a sufficiently large number of the character judges on one or more character traits.

THE TEST OF SUGGESTIBILITY

The test of suggestibility employed was one recently devised by Professor Hull in connection with an extensive research program involving both waking and hypnotic suggestibility. It has been described elsewhere.[2] The subject stands before the experimenter blindfolded. Quite without the subject's knowledge, a thread has been attached to her outer garment at the shoulder. This thread communicates to a light-running device which traces in detail on smoked kymograph paper the forward-backward postural movements of the subject. The subject is permitted to stand quietly thus for two minutes during

[2] C. L. Hull, Quantitative methods of investigating waking suggestion, J. Abn. and Soc. Psychol., 1929, Vol. 24, pp. 153–169.

which her spontaneous postural tremors are traced as a control record. Then the direct suggestion is given. It consists of telling the subject over and over in a quiet, confident tone that she is swaying forward. This is continued without interruption for approximately five minutes or until the subject sways so far forward that she loses her balance and must be caught to prevent stumbling or falling. This is called a "fall."

The experimenter was a graduate nurse and in addition had the status of instructor in the training school where she did supervising in nursing procedures. This gave her prestige among the pupils and doubtless greatly aided her in securing the cheerful cooperation of the subjects. Only one or two expressed unwillingness to go to the psychological laboratory to perform the experiment. This matter of prestige should also be considered in judging the responses of the subjects in the test. It may be added in this connection that the personality of the experimenter is somewhat milder than average, being distinctly persuasive rather than aggressive or commanding. Her manner while quiet and dignified sometimes shows a slight hesitancy which suggests a trace of something akin to timidity. Certain observations indicated, however, that even with subjects quite unacquainted with her, the quality of her voice and general method of administering the suggestions was above the average in effectiveness.

After the subject had arrived at the laboratory she was permitted to sit a few minutes in a waiting room. She was then blindfolded and led into the experimental room where she was placed with her feet on marks drawn on the floor at a convenient distance from the recording apparatus. The subject was told to make herself as comfortable as possible, that she would be harmed in no way and to relax and feel at ease. She was then told that she would be given some suggestions which she was to try neither consciously to follow nor deliberately to resist, but simply to relax and do what she felt like doing.

After giving these instructions, the kymograph was started and the subject stood quietly for approximately two minutes which gave the neutral control record. Then the experimenter,

standing a short arm's length before her, gave the postural suggestion in a quiet, confident tone, as follows:

You are falling forward, you are falling forward, forward. You are falling forward, falling forward, falling forward. You are coming over just a little, just a little, falling forward, falling forward, forward. Over a little farther. Forward, forward, farther forward. You are coming up on your toes now, up on your toes, up on your toes, now. You are falling over, falling over, coming over on your toes. Over, over. You are falling 'way over, 'way over, 'way over. Forward, falling forward, 'way over. You can't help yourself. You are coming 'way over, you can't help yourself. You are falling 'way over, forward, 'way over, falling forward, over, over. A little farther forward. That's it, 'way over. Forward, forward. You are almost touching me, almost touching me. I am right here to catch you, I won't let you fall. 'Way over, falling forward, over, over, 'way over, 'way over—

These suggestions were continued for a maximum of five minutes or until the subject "fell" forward. In such cases if the subject did not catch herself and resume her original position she was returned to it by gentle pressure applied at the shoulders. About a minute for recovery was permitted. Then the subject was again given the same suggestion exactly as before until the subject fell forward or until the maximum time had elapsed. A minute or so was allowed for the recovery from this suggestion, when the subject was directed to whisper aloud to herself the same suggestions as previously given by the experimenter: "I am falling forward, etc." As in the two preceding suggestions this continued for about five minutes or until the subject "fell." Then the subject was led from the room and the blindfold removed. She was given pencil and paper and asked to write out in detail her impressions during the experiment. Before being excused she was cautioned about talking of the experiment to any one lest it influence the other subjects and thus spoil the investigation.

THE CHARACTER DETERMINATIONS

Five persons made independent character judgments of the subjects tested for suggestibility. These character judges were women of superior general ability with nurse's training whose

duty it was to supervise the work of the pupil nurses. As a consequence every supervisor came into fairly close contact with every pupil nurse. Unfortunately the relationship of superiors to subordinates in hospitals is somewhat militaristic in its nature. This probably prevented the supervisors from obtaining as intimate insight into certain phases of the characters of the subjects as would be desirable. It must be added also that the supervisors had very little technical training in modern psychology. On the positive side it may be said that a supervisor in her professional contact with the pupil nurse has an opportunity to judge the latter's sympathy from her attitude towards her patients. The nurse's integrity is revealed by the extent to which she may be trusted to carry out vitally important orders. Her optimism is tested by the adverse conditions often encountered in hospital work. The sweetness of her temper is tried by the many unpleasant tasks assigned. Her tact is revealed by the way she manages to minimize the irritabilities of the patients whose nerves are worn by suffering.

As a preliminary to the making of the character judgments, the experimenter approached the supervisors individually and explained the procedure of the ranking. Since optimism was the first trait to be judged this was made the basis of the explanation. There were accordingly laid before the supervisor three cards each 3 inches by 5 inches in size upon which were lettered plainly the names and brief descriptions of three degrees of the trait in question. The supervisor was then given a pack of 70 cards about the size of ordinary calling cards upon each of which was written the name of one of the pupil-nurse subjects to be ranked. The supervisor was instructed first to sort out the 70 cards into three piles corresponding to the three degrees of the trait being judged as indicated by the three larger cards. The judge was then directed to arrange or rank each pile of cards in a vertical row, the card of that subject possessing the highest positive degree of the trait within a given pile being put at the top and the one possessing it least at the bottom. In this way three ranked series were produced, one for each pile of the cards. The supervisor was next instructed to join the

three sets of ranked cards appropriately into a single ranked series. Lastly the entire series thus produced was gone over as a whole for final judgment, particularly at the junctions of the series from the separate piles, and cards here and there rearranged so that they represented a final judgment of progressive decrement in the trait from top to bottom of the series.

In order to secure strictly independent judgments the supervisors were cautioned not to consult with each other or with any one about the personalities of the pupil nurses until the ranking should be completed. With a view to minimizing the "aura" effect the rankings of the several traits by a given judge were separated from each other by at least a day.

The traits ranked are given below in the order judged, together with the descriptive material placed on each of the three cards indicating degrees of the trait.

I. Optimism and pessimism.
   1st (card). Optimism—looking on the bright side of things.
   2nd (card). Average.
   3rd (card). Pessimism—disposition to look on the dark side of things.

II. Positive suggestibility and negativism.
   1. Suggestible—one who takes suggestion very easily, who acts just as you wish or expect, the most suggestible using no effort on their part, following the suggestion of another just as if it were her own.
   2. Average.
   3. Negativistic—one who acts oppositely, contradictorily, the reverse of that which we wish or expect; the opposite of suggestible.

III. Sympathy.
   1. Sympathetic—one who has a feeling of compassion for another person's suffering, the quality of being affected by the state of another with feeling corresponding in kind.
   2. Average.
   3. Unsympathetic—one who does not possess a feeling of compassion for another's suffering.

IV. Temper.
   1. Sweet-tempered—one who has a composure and calmness of mind, one who has the natural tendency to take everything, good or bad, with sweetness and calmness.
   2. Average.

3. Ill-tempered—one who is easily upset, easily irritated, has a tendency to take things with bitterness and irritability.

V. Personal beauty.
1. Pretty—perfection of form and features, pleasing to the eye.
2. Average.
3. Homely—one who is not pleasing to look at, imperfect of face or form.

VI. Tact in handling patients.
1. Tactful—one possessing an intuitive feeling of what is fit, best to be done, most proper. Ability to get the patient to do things her way and at the same time make the patient think it is the best and only way.
2. Average.
3. Tactless—one who does not know what is best to be done, one who irritates another person—"rubbing a person the wrong way."

VII. Honesty and sincerity.
1. Honest and sincere—being honest, trustworthy, true, just, upright in one's actions and profession, intending precisely what one says or what one appears to intend.
2. Average.
3. Dishonest and insincere—being dishonest, untrue, untrustworthy, unjust in one's actions and profession, not intending precisely what one says or what one appears to intend.

VIII. Positive suggestibility and negativism (repeated).

As pointed out above, positive and negative suggestibility was the trait of major interest. It was therefore placed second in the series to insure that the judges would be thoroughly familiar with the technique of ranking before attempting to judge it. It appears again at the end of the series for a second independent ranking so that a consistency coefficient might be obtained on the ranking of the group of judges as a whole. Personal beauty was inserted into the series as a kind of control on the assumption that it is a relatively objective trait and presumably not very closely related to the character traits as judged and hardly at all related to the behavior in the test.

RESULTS

The test

The amplitudes of the postural movements of the subjects were recorded on smoked paper by the apparatus employed at

one-third their actual amounts. There were in addition a time line recording 5-second intervals and a line for signals. The latter was used particularly to mark the beginning of the suggestion periods. Each suggestion record was scored in two ways—for amplitude of forward postural movement and for time, thus giving each subject six test scores. Where the subject "fell" there was little difficulty in scoring. The time score

TABLE 1

The means, probable errors of means and correlation coefficients of the six scores of direct suggestibility, 56 pupil nurses

	MEANS	PROBABLE ERROR OF MEANS	CORRELATION COEFICIENTS					
			Time to maximal reaction in seconds			Amplitude of maximal reaction in millimeters		
			First hetero-suggestion	Second hetero-suggestion	Auto-suggestion	First hetero-suggestion	Second hetero-suggestion	Auto-suggestion
Time:								
First hetero-suggestion	166.2	13.23						
Second hetero-suggestion	122.1	12.17	+.42					
Auto-suggestion	76.8	6.35	+.05	+.35				
Amplitude:								
First hetero-suggestion	61.9	5.02	.00	−.14	−.29			
Second hetero-suggestion	68.6	5.39	−.05	−.10	−.20	+.78		
Auto-suggestion	66.9	5.06	−.13	−.19	−.23	+.74	+.87	

was the number of seconds from the beginning of the suggestion to the fall. Since the amplitude of the "fall" varied according to how quickly E interrupted it by pressure from her hands, all such cases were given a flat value of 150 mm. In those cases where subjects showed no response whatever, the time was given as a flat 300 seconds even though the experimenter may have prolonged the suggestion beyond this time. Where the response was clear but less than a "fall," the amplitude was measured at

its highest point from a line beneath drawn at the level of the control tracing taken previous to the beginning of the suggestion. In some cases it was difficult to tell whether a slight rise in the curve was a spontaneous postural movement or a slight response to suggestion. On such occasions the scorer had no alternative save to exercise his judgment from a close examination of the normal control tracing and the record as a whole.

The means, probable errors and intercorrelations of the six test scores are given in table 1. There is no statistically significant difference in the amplitude of the reactions in the three tests. The second suggestion is the largest of the three. Fifteen subjects showed complete falls on this test, whereas, only 12 fell on the first and third trials. The time means show a very marked and statistically reliable decrease with the successive repetitions of the suggestion. The fall in duration from the first to the second is almost exactly the same as from the second to the third. The behavior under auto-suggestion does not appear to differ significantly from the other two. The progressive diminution in suggestion reaction time with this test has been observed repeatedly[3] though never before on such a large scale and with subjects entirely in the normal state.

Passing to the interrelations among the several measures of suggestibility in this experiment, we note that the nine coefficients between the time measures on the one hand and the amplitude measures on the other, are all either negative or zero in value. This is exactly as should be expected though it is perhaps a little surprising that the coefficients are so small. It should be especially noted that the coefficients quite uniformly grow larger from left to right in all three rows. We shall have occasion to refer to this later.

We next consider the intercorrelation among the three time measures. We have already seen reason to doubt that the psychology of these measures differs materially from one to the other. Accordingly these intercorrelations may be regarded

[3] Hull, *op. cit.*, pp. 158–159; also C. L. Hull and B. Huse, Comparative suggestibility in the trance and waking states, Am. Jour. Psychol., 1930, 42, pp. 279–286.

somewhat in the light of reliability coefficients. They are, .42, .05, and .35. As reliability coefficients these are exceedingly low. Passing to the three intercorrelations among the three amplitude measures we find .78, .74 and .87—a comparatively gratifying situation as regards reliability. This indicates that with subjects who show all degrees of suggestibility, amplitude of response is a much more promising score than is time of the maximal response. Of the three traits, the third (auto-suggestion, appears to show the highest promise. The first and third correlate better with the second than with each other.

TABLE 2

The reliability values for trait rankings by five judges of 56 pupil nurses

TRAITS	CORRELATION OF COMPOSITE OF TWO JUDGES WITH THAT OF THREE OTHER JUDGES	RELIABILITY COEFFICIENT OF ALL FIVE JUDGES BY THE FORMULA $\frac{2r}{1+r}$
Beauty	.39	.56
Optimism	.57	.72
Honesty	.61	.76
Sympathy	.23	.37
Sweet temper	.59	.74
Tactfulness	.57	.72
Suggestibility I	.30	.47
Suggestibility II	.29	.45

The traits

The rankings obtained from each of the five character judges were converted into equivalent linear units on a 100-point scale by a method now so well known that it need not be described.[4] After conversion the score of the subject possessing the most of the trait is a large figure, in the neighborhood of 100, the score of the median person is near 50 and the score of the person judged as possessing the least of the trait is small, being around zero.

[4] C. L. Hull, Aptitude testing, p. 386ff.

Since there were five independent rankings of each trait, it is possible to secure reliability coefficients which will give us an indication of the accuracy with which the rankings measure whatever is being measured by them. The experimenter divided the five supervisors into two groups, the first group being made up of two judges who were regarded as the best and the other group of the three remaining ones. The value obtained by combining the scores from the first group of judges was correlated with the corresponding value obtained from the

TABLE 3

Intercorrelation among character traits of 56 pupil nurses on the basis of composite judgments of five supervisors

	BEAUTY	OPTIMISM	HONESTY	SYMPATHY	SWEET TEMPER	TACTFULNESS	SUGGESTIBILITY
Optimism	+.41						
Honesty	−.03	+.45					
Sympathy	+.28	+.58	+.72				
Sweet temper	+.41	+.68	+.78	+.84			
Tactfulness	+.41	+.66	+.73	+.80	+.90		
Suggestibility I	+.21	+.48	+.45	+.65	+.75	+.75	
Suggestibility II	+.58	+.35	+.17	+.49	+.53	+.49	+.55

scores by the second group of judges for each trait. The result in each case was then substituted in the formula $\frac{2r}{1+r}$ which yields the probable reliability coefficient of all five judges when combined. These values are given in table 2. They show that sympathy is the least reliably judged. It is a little disappointing that suggestibility, the trait of chief interest, is next worst judged, the corrected coefficient being less than .50 in both cases. Beauty, supposedly an objective trait, fares little better. The remaining traits are relatively satisfactory, all being above .70.

The intercorrelations among the traits are shown in table 3. Several coefficients there displayed are of interest. First may be mentioned the curious negative correlation between beauty

and honesty. The "aura" effect would naturally tend to produce a positive correlation here even if there were none in reality. This −.03 may indicate a fair sized tendency for plain girls to be more sincere than the more handsome ones. The size of the coefficients indicate that tactfulness, sweet temper and suggestibility form a very homogeneous group, at least from the point of view of the five supervisors of nurses.

The two determinations of suggestibility as a trait of character yield a set of coefficients of special interest. It will be recalled that the first was secured near the beginning of the rating process whereas the second was made as the eighth and last of the series of trait rankings. We have already seen that the reliabilities of the two were almost exactly alike, .47 and .45. It is apparent, nevertheless, that the second rating yields a somewhat different result from the first. This is shown by the consistently smaller correlations with the other traits yielded by the result of the second ranking. Since the reliabilities of both are the same, the indication is that through their practice in judging and ranking character traits the judges have become more discriminating and as a consequence have eliminated part of the aura effect. Lastly there may be mentioned the correlation between the two independent determinations of suggestibility, +.55. This might be considered a reliability coefficient for the measurement of suggestibility if it were assured that the substance of the trait ranked by the judges had not changed in the interval between the two rankings. This coefficient is some ten points higher than the reliability coefficients determined from the separate rankings by formula.

Validity of the test

The answer to our primary question as to whether or not direct suggestibility is a desirable trait and, in particular, whether it has any relationship to the traits called suggestibility in everyday life, is found in table 4. In studying this table it should be borne in mind that since high suggestibility by the test gives a *small* time score and a *large* amplitude score, the correlations for time would be expected by our hypothesis to be

negative, whereas those with amplitude would be positive. A glance at the signs of that part of the table below "optimism" shows that this expectation is substantially realized. On the right half of the table showing the correlations with amplitude, there is not a single exception to the expectation. This may possibly be due to the distinctly higher indications of reliability of this measure. On the time side the picture is fairly consistent although there are exceptions particularly in the column for

TABLE 4

Table showing the correlations of the measures of direct suggestibility with the several ranked traits

TRAITS	MEASURES OF DIRECT SUGGESTIBILITY					
	Time to maximal reaction in seconds			Amplitude of maximal reaction in millimeters		
	Hetero-suggestion	Hetero-suggestion	Auto-suggestion	Hetero-suggestion	Hetero-suggestion	Auto-suggestion
Beauty	−.11	−.24	+.04	−.09	−.04	+.05
Optimism	+.06	−.14	−.06	−.07	−.24	−.16
Honesty	+.10	+.03	−.28	+.15	+.08	+.09
Sympathy	+.16	−.03	−.21	+.08	+.03	+.11
Sweet temper	+.03	−.06	−.19	+.11	+.02	+.08
Tactfulness	+.02	−.19	−.27	+.12	+.02	+.07
Suggestibility I	−.01	−.17	−.19	+.22	+.08	+.18
Suggestibility II	−.02	−.07	+.06	+.18	+.16	+.28
Mean of six lower rows	+.05	−.08	−.18	+.14	+.07	+.14

first hetero-suggestion. In the second column the tendency is marked, there being but one small exception out of six. In the third column the tendency is still more marked, particularly when one considers the size of the coefficients. This curious tendency for the coefficients to become more negative from left to right in the time measure we noted above when we considered table 1. It probably reflects some characteristic difference between the determining factors involved in the three repeti-

tions of the test. The positive correlation with the time of simple hetero-suggestion is almost as consistent as the negative correlations of the other two.

From the point of view of the traits it may be observed that as regards amplitude, optimism shows a consistent reversal of the sign of the coefficient. Perhaps honesty also should be placed in the doubtful class since it has two exceptions as to the signs expected. None of the others, however have more than one exception. Sympathy, sweet temper, tactfulness and suggestibility seem to form a fairly homogeneous group with little to distinguish one from the others. It may be significant that suggestibility comes off with the most perfect agreement of signs with expectation, there being but a single exception in twelve coefficients.

The size of the coefficients in table 4, is characteristically and uniformly small. It is certain from the size of the reliability coefficients that this is due in part to attenuation from the crudities of measurement. Naturally the probable errors of such small values are rather large ranging around $\pm.09$. One or two isolated coefficients of this size would mean nothing. But the striking consistency of the 36 coefficients below "optimism" is distinctly convincing. This is particularly true if honesty be excluded from the series. The intercorrelations among these traits run approximately as high as their reliability coefficients which indicates that the substance behind the ratings was about the same in all these cases, though differently named. It is difficult to characterize this with any certainty but it evidently was something regarded as desirable in a person's character. The distinctly clear showing on suggestibility together with the *a priori* probability that tactfulness, sweet temper and sympathy at least are dependent to a considerable extent on suggestibility, indicates that the postural suggestion test really is a test of suggestibility in the sense of a character trait.

The above results are of interest from several points of view. In the first place it is one of the very few tests designed to test a character trait which has ever been shown to have any real

correlation with the character trait in question. Secondly it indicates quite definitely that suggestibility, instead of being an undesirable trait is in reality a desirable one with some indication of being a component of several eminently desirable traits. In view of the foregoing considerations it would seem worthwhile to try out this test along with others in the derivation of aptitude test batteries for vocations in which character traits resembling those listed above appear upon preliminary analysis to be important for success.

THE THEORY OF TWO FACTORS: AN ALTERNATIVE EXPLANATION*

PART I. GENERAL FACTOR, "G"

HENRY F. ADAMS

University of Michigan

That tables showing the intercorrelations of mental tests are prone to assume an hierarchical form was announced by Spearman[1] as early as 1904. Described mathematically, the various entries in the table must satisfy the generalized formula

$$r_{ap}/r_{aq} = r_{bp}/r_{bq},$$

which is obviously a proportion. Solving, it becomes

$$r_{ap} \times r_{bq} = r_{aq} \times r_{bp},$$

or

$$r_{ap} \times r_{bq} - r_{aq} \times r_{bq} = 0.$$

The last form has been called the tetrad equation, and the result obtained by solving it the tetrad difference. Similarly, a table of intercorrelations between tests or any other measurements which satisfies the tetrad equation may be said to have a tetrad form.

Whenever a series of values varies as regularly and dependently as in the tetrad form of table, it is impossible not to believe in some causal factor or factors which are responsible for the proportionate variation. For Spearman the correlation between two series "may be forthwith assumed to indicate and

* This investigation was made possible through a grant from the Faculty Research Fund.

[1] American Journal of Psychology, 1904, vol. 15, pages 202–293.

measure *something common to* both series in question."[2] If the common factor runs through more than two series, but in varying amounts, the ensuing table of correlations should assume the tetrad form. When this happens, and only when it does, each measurement included in the table of correlations can be divided into two parts: one a general factor called g, and the other a specific factor called s. The g has recently been identified with mental energy, and s with the cortical engines. The correlation between a certain test a and g may be determined, provided at least three different tests have been given, by the generalized formula

$$r_{ag} = \sqrt{r_{ab} \cdot r_{ac}/r_{bc}} = \sqrt{r_{ac} \cdot r_{ad}/r_{cd}}, \text{ etc.}$$

The purpose of this paper is to show the possibility of substituting for the Spearman g, or general factor, another term, x, which represents the *standard* in terms of which the performance is evaluated. Mental test scores are ordinarily given as functions of the number of items correctly done divided by the total number of items. To determine whether an answer is correct, it is checked against the *standard* furnished by the maker of the test. A perfect performance consequently coincides with the standard. Thus, whenever a test is scored in terms of a standard, it is possible to correlate each performance against the standard and average for the group. This coefficient showing the average correlation between the individual performances and the standard may be represented by the expression r_{ax}. Similarly, the correlation between any other test and its standard may be given the form r_{bx}. It will be shown that x may be substituted for g in all of Spearman's more fundamental formulas without doing violence to the algebra involved.

The alternative explanation of the tetrad table, being based upon the idea of an external or objective standard, is not com-

[2] American Journal of Psychology, 1904, Vol. 15, page 258. The same point is made by a number of other writers. Among them are Brown and Thomson, Essentials of Mental Measurement, page 176; Hull, Journal of Educational Research, May, 1927, page 1; Kelley, Cross Roads in the Minds of Men.

mitted solely to a program of intelligence testing, and has consequently a far wider range of application than the Spearman theory. It accounts satisfactorily for Thomson's hierarchy without a general factor; it explains the tetrad form of table obtained by Wilson showing the correlations between the precipitations reported by six stations in New England; it even accounts for the tetrad form of the table showing the probable errors of coefficients of correlation as affected both by the size of r and the number of cases. It is general in its application rather than limited to one particular phase. As such, it is a more useful and desirable hypothesis.

The alternative explanation depends upon the validity of a fundamental formula, which is that the correlation between two tests is the product obtained by multiplying the correlation of test a with its standard by the correlation of test b with its standard. The algebraic expression is

$$r_{ab} = r_{ax} \cdot r_{bx}$$

This equation can hold only when the two x's are equal and when the standards are identical in *form*.

In the Spearman terminology, this formula becomes

$$r_{ab} = r_{ag} \cdot r_{bg},$$

and this formula has already been proved by Garnett.[3] His whole argument should be read for an understanding of the limiting conditions, but the essential proof is quite simple. According to Spearman,

$$r_{ag} = \sqrt{r_{ab} \cdot r_{ac}/r_{bc}},$$

and

$$r_{bg} = \sqrt{r_{ba} \cdot r_{bc}/r_{ac}}.$$

Multiplying,

$$r_{ag} \cdot r_{bg} = \sqrt{r_{ab} \cdot r_{ac} \cdot r_{ab} \cdot r_{bc}/r_{bc} \cdot r_{ac}}$$

[3] Garnett, J. C. M., Proceedings of the Royal Society, Series A, 1919, Vol. 96, pages 91–111. See especially page 101.

By cancellation,

$$r_{ag} \cdot r_{bg} = r_{ab}.$$

If x and g are equal, x may be substituted for g in these formulas, and $r_{ab} = r_{ax} \cdot r_{bx}$.

The idea expressed by the last formula was derived from certain relations revealed by the analysis of data obtained from discrimination experiments. Since an account of the materials and techniques has already been published,[4] a full account will not be given here. In all of the experiments the rank difference method was used. Series of ten items were arranged in order of some attribute on two separate occasions by each subject. The discriminations included weight, length, and area, and consequently the true or standard order for each series was ascertainable. Each arrangement made by each subject could therefore be correlated with the standard or correct order, and the average value obtained. This gave the values r_{ax}, r_{bx}, r_{cx}, ..., r_{nx} for the different series.

Each subject, furthermore, performed each experiment twice. His first order could therefore be correlated with his second order, and the average for the group taken. Thus a second series of values, r_{aa}, r_{bb}, r_{cc}, ..., r_{nn}, were obtained.

Comparisons were then made between the values of r_{ax} and r_{aa}, r_{bx} and r_{bb}, ..., r_{nx} and r_{nn}. It was found that $r_{aa} = r_{ax} \cdot r_{ax}$, $r_{bb} = r_{bx} \cdot r_{bx}$, and so on for the rest of the available material. The values r_{ax} and r_{ax}, r_{bx} and r_{bx}, etc., were so close together that the formula may be written, with a high degree of precision in the more useful and convenient form $r_{aa} = r^2_{ax}$. That is to say, the self correlations, or so-called "reliabilities," equal the square of the correlation of the test with its standard.

The discovery of this fundamental relation suggested that a table be made showing the intercorrelations of the different discrimination tests in which the same subjects had been used

[4] Adams, H. F., Journal of Social Psychology, 1930, vol. 1, pages 122–135.

throughout. Four sets of material were available. The first table shows the intercorrelations of five series of discriminations of area and five series of discriminations of length. Seventy subjects had performed all ten experiments. The correlation of each test with its standard is shown on the diagonal of the table; the empirical correlations appear in the customary coordinate positions. The values obtained by the use of the formula $r_{ab} = r_{ax} \cdot r_{bx}$ are given in parentheses under the cor-

TABLE 1

	C_1	C_2	C_3	C_4	C_5	L_1	L_2	L_3	L_4	L_5
C_1	.96									
C_2	.90	.94								
	(.90)									
C_3	.88	.85	.92							
	(.88)	(.86)								
C_4	.82	.81	.78	.85						
	(.82)	(.80)	(.78)							
C_5	.77	.75	.73	.69	.80					
	(.77)	(.75)	(.74)	(.68)						
L_1	.89	.87	.84	.80	.74	.93				
	(.89)	(.87)	(.85)	(.79)	(.74)					
L_2	.82	.82	.78	.74	.69	.79	.85			
	(.82)	(.80)	(.78)	(.72)	(.68)	(.79)				
L_3	.75	.73	.71	.67	.62	.72	.66	.78		
	(.75)	(.73)	(.72)	(.66)	(.62)	(.73)	(.66)			
L_4	.66	.64	.63	.60	.57	.65	.60	.57	.69	
	(.66)	(.65)	(.63)	(.59)	(.55)	(.64)	(.59)	(.54)		
L_5	.61	.60	.57	.55	.52	.60	.58	.50	.47	.64
	(.61)	(.60)	(.59)	(.54)	(.51)	(.60)	(.54)	(.50)	(.44)	

responding empirically obtained values. The evidence is presented in table 1.

The mean discrepancy between the theoretically and the empirically obtained values of r_{ab} is 0.0076, with an average probable error of 0.03. No single discrepancy is greater than the probable error of the coefficient involved.

In published works, however, it is ordinarily impossible to find given the correlations of each test with the standard.

Since the intercorrelations of the tests are presented, the r_{ax}, r_{bx}, ... , r_{nx} values can be computed algebraically. The r's may be omitted from the formula, for each subscript represents a pure number. The formula is determined as follows.

$$ax \cdot bx = ab$$

$$ax \cdot cx = ac$$

Then

$$ax = ab/bx, \text{ and } ax = ac/cx.$$

Multiplying,

$$ax \cdot ax = ab \cdot ac/bx \cdot cx$$

But

$$bx \cdot cx = bc.$$

Substituting, the formula becomes

$$ax \cdot ax = ab \cdot ac/bc,$$

or

$$ax = \sqrt{ab \cdot ac/bc}$$

A comparison between the theoretically and the empirically determined values of r_{ax} in table 1 follows.

	EMPIRICAL	THEORETICAL	DIFFERENCE
C_1	.96	.96	.00
C_2	.94	.94	.00
C_3	.92	.92	.00
C_4	.85	.86	.01
C_5	.80	.79	.01
L_1	.93	.92	.01
L_2	.85	.85	.00
L_3	.78	.78	.00
L_4	.69	.72	.03
L_5	.64	.63	.01
Mean difference			.007

The two methods give almost identical results, the average difference being well inside of the probable error. In this situation at least it follows that $r_{ab} = r_{ax} \cdot r_{bx}$ where x indicates the standard or perfect performance.

The second bit of evidence was also obtained from discrimination experiments in which size and shape were varied simultaneously. The judgments were in terms of size. Fifty subjects were used in each of eight experiments. The results

TABLE 2

	1	2	3	4	5	6	7	8
1	.99							
2	.97	.98						
	(.97)							
3	.82	.81	.82					
	(.81)	(.80)						
4	.79	.79	.67	.82				
	(.81)	(.80)	(.67)					
5	.85	.84	.74	.72	.86			
	(.85)	(.84)	(.71)	(.71)				
6	.97	.97	.80	.79	.85	.99		
	(.98)	(.97)	(.81)	(.81)	(.85)			
7	.91	.89	.75	.74	.79	.90	.90	
	(.89)	(.88)	(.74)	(.74)	(.77)	(.89)		
8	.90	.90	.77	.72	.78	.90	.85	.91
	(.90)	(.89)	(.75)	(.75)	(.78)	(.90)	(.82)	

are given in table 2. As before, the r_{ax} values appear along the diagonal of the table, the empirical coefficients in the coordinate positions, and the theoretical in parentheses below the empirical.

Here, again, the evidence is in support of the contention that $r_{ab} = r_{ax} \cdot r_{bx}$, the average discrepancy being 0.01 ± 0.03.

A comparison of the empirical and the theoretical r_{ax} values, the latter being obtained by the use of the formula $r_{ax} = \sqrt{r_{ab} \cdot r_{ac} / r_{bc}}$ follows.

	EMPIRICAL	THEORETICAL	DIFFERENCE
1	.99	.99	.00
2	.98	.98	.00
3	.82	.83	.01
4	.82	.81	.01
5	.86	.87	.01
6	.99	.98	.01
7	.90	.91	.01
8	.91	.91	.00
Mean difference			.0063

The differences are negligible, and emphasize again the validity of the assumption.

The third piece of evidence was obtained from six series of discrimination experiments in which two or more simultaneous variables were present. The discriminations were made in

TABLE 3

	1	2	3	4	5	6
1	.21					
2	.21	.98				
	(.21)					
3	.21	.96	.98			
	(.21)	(.96)				
4	.21	.96	.96	.98		
	(.21)	(.96)	(.96)			
5	.22	.96	.96	.96	.98	
	(.21)	(.96)	(.96)	(.96)		
6	.19	.94	.94	.93	.93	.96
	(.20)	(.94)	(.94)	(.94)	(.94)	

terms of the relative size of squares which contained various numbers of colored dots. In experiments 2–6 inclusive the areas of the squares and the number of dots contained therein were correlated by +1.00, +0.50, 0.00, −0.50, and −1.00. In

series 1, the squares were all of the same size, but the number of dots varied from square to square. Fifty subjects performed the experiment twice. The results are given in table 3. The r_{ax} values follow the diagonal of the table, the empirical r_{ab} values are placed at the coordinate points, and the theoretical r_{ab} values appear in parentheses under the corresponding empirical value.

The discrepancy between the empirical and theoretical values of r_{ab} is 0.003 ± 0.02. When the r_{ax} values are computed by the use of the formula given above, the discrepancy between the empirical and the theoretical values is zero.

TABLE 4

	1	2	3	4	5
1	.92				
2	.79	.83			
	(.76)				
3	.41	.28	.47		
	(.43)	(.39)			
4	.55	.53	.40	.62	
	(.57)	(.51)	(.29)		
5	.51	.43	.24	.38	.58
	(.53)	(.48)	(.27)	(.36)	

The fourth piece of evidence was obtained from an experiment involving the discrimination of the areas of circles. There were five series of ten circles each, exactly like those reported in table 1, only in this experiment all of the ten circles in each series were drawn upon one large sheet of paper. Hence there was every chance for illusions of size to appear because of contrast and confluxion. The set up of the experiment was unfavorable to bring about proportionate changes in the correlational values. The results are given in table 4.

For the fourth time the evidence is pleasingly harmonious, the mean discrepancy being 0.043 ± 0.085. The comparison of the theoretical and empirical values of r_{ax} is as follows:

	EMPIRICAL	THEORETICAL	DIFFERENCE
1	.92	.92	.00
2	.83	.84	.01
3	.47	.44	.03
4	.62	.61	.01
5	.58	.58	.00
Mean difference.....			.01

From the evidence presented two important conclusions may be drawn. First, the formula $r_{ab} = r_{ax} \cdot r_{bx}$ is justified empirically. In the 98 cases cited the discrepancy between the empirical and the theoretical values is on the average below the probable error of the coefficients involved. The maximum difference is less than $1\frac{1}{2}$ times the probable error. Under the conditions of the experiments the correlation between any two series is the product obtained by multiplying the correlation of one series with its standard by the correlation of the other series with its standard when they are identical in form.

In the second place it has been shown that the r_{ax} values, if unknown or not given, can be computed with a high degree of precision by the use of the formula $r_{ax} = \sqrt{r_{ab} \cdot r_{ac} / r_{bc}}$. In the 29 cases given the discrepancy between the empirical and theoretical values is considerably less than the probable error of the coefficients involved.

If we wish to analyze our material by the Spearman technique we must first determine the r_{ag} values. To do this we must use the formula $r_{ag} = \sqrt{r_{ab} \cdot r_{ac} / r_{bc}}$. But we have just shown that this is the formula for computing the r_{ax} values. Consequently as far as this evidence is concerned what Spearman would call the saturation of the test with g becomes the correlation of the performance with the standard in terms of which the performance is evaluated. The last clause might be altered to read the saturation of the group with knowledge of the standard and their ability to apply such knowledge to the practical situations brought out by the test.

It has been said that the two fundamental formulas may be

expected to hold only under the conditions of the experiments described, for it was these results which led to the development of the principles. In all cases the rank difference technique was employed. The number of items arranged in order of merit was invariably 10, which means that the standard deviations were equal. The groups used were relatively homogeneous in age, there being less than a two-year spread. They were not homogeneous as regards sex but no significant sex differences appeared. Most important the standards were identical in form throughout, every judgment being based upon the relative magnitude of the items. It makes no difference, consequently, whether the particular experiments demanded the judgment of the relative magnitude of the areas of circles, or the relative magnitude of lengths of lines, or the relative magnitude of the number of dots. Magnitude or quantity was the standard and the standard was identical in all of the experiments.

A clear comprehension of the importance of the rôle played by the standard is so essential that a few illustrations will be given. Let us take the experiments in which the two simultaneous variables of size and number were used. If the ten items are arranged in order in terms of size at one time and later arranged in order of the number of dots, the correlations can be determined in either of two ways. On the one hand we can keep the items A, B, C, ..., J constant, recording first their relative size, and second the relative number of dots each contains. The result will then depend upon the set up of the experimental material. If the smallest area contains the smallest number of dots and the largest the greatest number of dots, and there is throughout a steady progression, then the correlation will approach $+1.00$, which will be the correlation between the standards. If the material is arranged in such a way that the smallest area contains the greatest number of dots, and the largest area the smallest number, then the correlation will approach -1.00, showing again the correlation of the standards. This method of computing correlation coefficients will give values which approach to the true correlation between standards, but will never equal it until random errors are

eliminated. This method then will give only the correlation between the two standards inherent in the same group of objects.

On the other hand, we can keep constant the magnitudes to be judged regardless of what items they are associated with, and then correlate the rankings for size magnitude with those for number magnitude, thus comparing the "ability" to make the two discriminations. In this case we appear to be correlating the two "abilities" rather than the standards, for in this case the standards are identical in form, being simply the numbers 1, 2, 3, ..., 10. It is only in this last case or in others analogous to it, that the two fundamental formulas may be expected to hold.

Furthermore, in the experiments already described the r_{ax} coefficients were computed by comparing the positions assigned each object with the true position of the object, and averaging for the group. Similarly, the r_{aa} coefficients were obtained by comparing the first order for each subject with the second order assigned by the same subject and averaging for the group. In other words the correlations were obtained as a result of item to item comparisons. And this is a conventional method of correlating such materials.

But this is no guarantee that the principles so far discovered will also apply to mental test correlations. With tests the technique is very different for the correlations are ordinarily determined by a comparison of end scores rather than by a comparison of items. It is, however, reasonable to believe that variations in end scores should be correlated with variations in performance. If this is the case the same principles may be inherent in the two methods of correlation. In fact, in so far as mental tests rank rather than measure the individuals in a group the same fundamental principles may well be involved.

To determine whether this is the case the following procedure was adopted. Tables giving the intercorrelations between mental tests were collected and the formula $r_{ax} = \sqrt{r_{ab} \cdot r_{ac} / r_{bc}}$ applied to give the r_{ax} values. Then, by multiplication of the $r_{ax}, r_{bx}, \ldots, r_{nx}$, values two by two, the theoretical $r_{ab}, r_{ac}, r_{bc}, \ldots, r_{yz}$ values were computed and compared

with the empirically obtained values. The mean discrepancy between the empirically and theoretically obtained values was obtained and compared with the mean probable error of the empirical coefficients in the table. The mental tests selected all had the common standard of per cent correct, or number of items correct out of a total number, or right minus wrong. It

TABLE 5

PAGE	DISCREPANCY	P.E.
145	.071	.074
147	.053	.075
147	.019	.021
148	.080	.080
149	.039	.046
Mean	.052	.059

TABLE 6

SOURCE	DISCREPANCY	P.E.
B. J. 13:395	.062	.070
B. J. 16:208	.035	.045
B. J. 17:36	.036	.039
B. J. 17:104	.056	.068
Psy. St. U. of L	.044	.047
B. J. M. S. No. 1	.033	.048
Mean	.044	.053

is presumably that these standards differ somewhat from each other, and certain consequent irregularities may be expected to appear. But the standards are sufficiently similar to yield first approximations. Correlation tables including any considerable number of tests without a standard, such as height, weight, rate of tapping, reaction time, etc., were excluded from consideration.

The first material selected for analysis consisted of the tables

presented by Spearman in support of his theory.[5] This evidence is presented in table 5. In the first column is the page bearing the table; in the second the average discrepancy between the theoretically and empirically obtained values; in the third the average values of the probable errors of the empirical coefficients.

The second set of tables was selected at random from those presented by writers who appear from context to favor the Spearman theory. All but one are taken from the British Journal of Psychology. The data obtained from the tables offered by these six writers are given in table 6.

TABLE 7

SOURCE	PAGE	DISCREPANCY	P.E.
Thorndike. Meas. of Int.	558	.031	.045
Viteles, J. App. Psy. 6.	395	.032	.045
Hollingworth, Voc. Psy.	172	.072	.080
Wilkins, Arch. Psy., No. 102.	23	.023	.034
Mean		.040	.056

A third set of tables was culled from writers who are not avowed supporters of the Spearman theory. Their evidence is presented in table 7.

The fourth set of material is in a way the most interesting exhibit to be offered, for the three tables selected have nothing whatever to do either with "ability" or intelligence as affecting the value of the intercorrelations.

The first is taken from Thomson's article, A Hierarchy Without a General Factor.[6] The experiment was one in dice casting. Since a maximum possible score can be obtained by multiplying the number of dice by 6, any individual cast can be expressed as a percentage of the maximum. Consequently a standard exists, and it may be expected that the fundamental formula

[5] The Abilities of Man, pages 145, 147, 148, 149.

[6] Brit. Jour. Psychol., 8: 272.

will apply. The necessary computations revealed a discrepancy between the theoretically and empirically obtained values of .035 with a probable error of .062 if 100 cases are assumed.

The second table was found in Wilson's discussion of precipitation as reported by six stations.[7] Here there is the common standard, inches of rainfall, running through all of the intercorrelations. In this table the discrepancy is .038 with a probable error of .046.

The third bit of evidence was obtained from the table given by Rugg[8] showing the probable errors of coefficients of correlation for different values of r and different numbers of measure-

TABLE 8

D/P.E.	FREQ.	D/P.E.	FREQ.
0.0–0.4	201	3.0–3.4	10
0.5–0.9	149	3.5–3.9	3
1.0–1.4	105	4.0–4.4	4
1.5–1.9	41	4.5–4.9	1
2.0–2.4	17	5.0–5.4	2
2.5–2.9	13		

N = 546.

ments. The common standard effective throughout the table is a mathematical formula. The first five columns of this table were treated as if they made a correlation table. When the technique was applied it was found that not only was the discrepancy zero but that the tetrad differences were also zero, approximately.

An alternative method of presenting the same evidence is to compare the discrepancy between the theoretically and empirically obtained values of the coefficients with the probable error of each empirically obtained coefficient. The comparison may take the form of the ratio, D/P.E. If this ratio is on the average 1.00 or less, and if the individual ratios do not exceed 5.0, and

[7] Science, N. S., Aug. 24, 1928.

[8] Statistical Methods, p. 404.

if the distribution is relatively normal, the evidence is very favorable to the theory. Table 8 shows the results of this method of dealing with the material. The two entries in excess of 5.0 are derived from the Spearman table on page 145.

The mean value of the ratios is 0.818 ± 0.022.

The last piece of evidence to be presented is taken from Spearman's original article.[9] In this quotation it seems to the writer that Spearman admits the main contention of the first part of this paper, namely, the rôle played by the standard. The quotation is as follows:

"A target was constructed of a great many horizontal bands, numbered from top to bottom. Then a man shot successively at a particular series of numbers in a particular order; clearly, the better the shot, the less numerical difference between any number hit and that aimed at; now, just as the measurement of any object is quite appropriately termed a 'shot' at its real value, so conversely, we may perfectly well consider the series of numbers actually hit in the light of a series of measurements of the numbers aimed at. When the same man again fired at the same series, he thereby obtained a new and independent (provided, of course, that there be no appreciable constant error) series of measurements of the same set of objects. Next, a woman had the same number of shots at some set numbers in a similar manner. If, then, our above reasoning and formulae are correct, it should be possible, by observing the numbers hit and working out their correlations, to ascertain the exact resemblance between the series aimed at by the man and woman respectively. In actual fact, the sets of numbers hit by the man turned out to correlate with those hit by the woman to the extent of 0.52; but it was noted that the man's sets correlated to one another to 0.74, and the woman's sets with one another to 0.36; hence the true correspondence between the set aimed at by the man and that aimed at by the woman was not the raw 0.52 but

$$\frac{0.52}{\sqrt{0.74 \times 0.36}} = 1.00$$

that is to say, the two persons had fired at exactly the same series of bands, which was really the case."

They had fired at exactly the same series of bands; they had used and approximated exactly the same standard.

Very strong and convincing evidence has been presented in

[9] Amer. Jour. Psychol., 1904, vol. 15, p. 271.

support of our fundamental formula $r_{ab} = r_{ax} \times r_{bx}$, and its derived companion

$$r_{ax} = \sqrt{\frac{r_{ab} \times r_{ac}}{r_{bc}}}.$$

They have been shown to hold, inside the limits imposed by the probable error, in over 600 cases without exception when the standards involved in the various intercorrelations were identical in form. This is very strong inductive proof. The validity of the formulas may be accepted as a working hypothesis pending strictly mathematical analysis if the reader is not satisfied with Garrett's demonstration. In the following arguments, we shall accept them as true provisionally.

Several important deductions can now be made. In presenting them the cumbersome r's will be eliminated and only the subscripts retained for algebraic treatment. It seems to be perfectly fair to do this, for the subscripts stand for definite numerical values. The further assumption is made that all errors are random and uncorrelated.

The first point to be made is that when the formula $ab = ax \times bx$ holds, the tetrad equation must be satisfied. The tetrad equation takes the form.

$$ap \times bq - aq \times bp = 0$$

Now

$$ap = ax \times px$$
$$bq = bx \times qx$$
$$aq = ax \times qx$$
$$bp = bx \times px$$

Substituting in the tetrad equation,

$$ax \times px \times bx \times qx - ax \times qx \times bx \times px = 0,$$

as must be the case, since all the terms in the equation cancel. Therefore, when $ab = ax \times bx$, the tetrad difference must equal

zero. Nothing is said about the converse of the proposition. For certain other forms of tetrad equation may be perfectly well be equal to zero.

In the second place, the formula for correction for attenuation due to random errors is just another way of expressing the fundamental formula.'

$$ab = ax \times bx = \sqrt{aa} \times \sqrt{bb}$$

Dividing both sides of the equation by the same quantity, it becomes

$$\frac{ab}{\sqrt{aa \times bb}} = \frac{\sqrt{aa \times bb}}{\sqrt{aa \times bb}} = 1.00$$

Correction for attenuation may be expected to give 1.00 when the standards are identical in form and the errors are uncorrelated. The ab values are therefore functions of the uncorrelated errors, and will shift up and down with the decrease and increase of the number of the errors.

In the third place, since $ab = \sqrt{aa \times bb}$, the self correlations of the tests must conform to the tetrad form of table. That they do not ordinarily thus conform is a very good sign of the presence of correlated errors which inflate the values.

In the fourth place, application of the formula for partial correlation shows that when the standard is partialed out, $ab = 0$.

Let r_{12} be represented by ab, which $= ax \times bx$

Let r_{13} be represented by ax

Let r_{23} be represented by bx.

Then the formula becomes.

$$r_{ab \cdot x} = \frac{(r_{ax} \times r_{bx}) - (r_{ax} \times r_{bx})}{\sqrt{(1 - r^2_{ax})}\,(\sqrt{1 - r^2_{bx}})} = 0$$

This means that a and b owe their correlation to the identity in form of the standard. When the standards coincide, the true correlation is at its maximum; when there is but chance agreement between the two standards, r_{ab} equals zero; when the standards are exactly opposite, r_{ab} has a maximum negative

value. This is equivalent to saying that the correlation ab is actually a product obtained by multiplying five terms, $a \times x \times b \times x \times y$, in which

a = the extent to which performance a approximates its standard.

x = the standard of a. It has the value of 1.

b = the extent to which performance b approximates its standard.

x = the standard of b. It has a value of 1.

y = the correlation between the two standards.

In practice, since x and x may be put equal to 1.00, $r_{ab} = ax \times bx \times y$. When y also equals 1.00, it becomes equal to x, and the correlation is $ax \times bx$. This point of view will be further developed in the section on Group Factors (to be published in April).

Now Spearman's r_{ag} and our r_{ax} are numerically equal, for both are equal to

$$\sqrt{\frac{r_{ab} \times r_{ac}}{r_{bc}}}$$

Consequently, his g must equal our x. Since x is identified with the standard, g must be also. If this is so, neither x nor g can be an index of "ability." On the other hand, a and b may be regarded as indices of "ability" if "ability" be thought of as the extent to which a performance approaches the standard. g and x, therefore, become limiting standards having values of 1.00 whenever the tetrad difference criterion is satisfied. If the "ability" is indicated only by a, and not by x, then division of the total "ability" into parts must be accomplished by an analysis of a. g and x disappear from the picture.

ON THE VALIDITY OF SILENT READING TESTS

M. EUSTACE BROOM, JOSEPHINE DOUGLAS,
AND MARION RUDD

State Teachers College, San Diego

There are numerous instruments available which are devised for the purpose of measuring the ability of pupils in the different grades in the elementary and in the junior high school to comprehend the meaning of the printed English language when they read it silently. Few detailed studies of the validity of these tests have been reported. By validity is meant the correspondence between the ability which is measured by a given test and the ability as objectively defined elsewhere and as measured by other similar tests. In the present study the coefficient of correlation is used to express the degree of relationship between the paired measures yielded by the different instruments. Each series of measures is correlated with each other series.

The data in table 1 were obtained in grades 7 and 8 in the Escondido, California, elementary school during the school year of 1928-1929.

The data in table 2 were obtained in grade 9 in the Woodrow Wilson Junior High School in San Diego during the school year of 1927–1928.

The data in table 3 were obtained in grades 8 and 9 in a private school in San Diego during the school year of 1927–1928.

The data in table 4 were obtained in grade 6 of the Chula Vista, California, Elementary School during the school year of 1928–1929.

From the mass of data presented, we may draw certain tentative conclusions:

1. Certain of the standardized tests seem highly valid, par-

ticularly the Sangren-Woody Reading Test; the Haggerty Reading Examination, Sigma 3; and the reading section of the Stanford Achievement Test, Advanced Examination.

TABLE 1

Validity coefficients for four silent reading tests and for one test of proof reading skills (The Briggs English Form Test)

Note: The probable errors for all coefficients except those involving Briggs English Form Test varied between .00 and .08; for those involving this test, between .07 and .15. Numbers of cases are shown in parentheses below the coefficients.

	MONROE (COMPREHENSION)	DETROIT	STANFORD READING	HAGGERTY SIGMA 3	BRIGGS
Monroe (Rate)	.79 (57)	.61 (44)	.39 (50)	.49 (45)	.01 (21)
Monroe (Comprehension)		.62 (43)	.42 (50)	.59 (46)	−.14 (21)
Detroit			.85 (75)	.99 (53)	.07 (49)
Stanford				.76 (48)	.50 (49)
Haggerty					.37 (22)

TABLE 2

Validity coefficients for three silent reading tests. 99 cases

Note: The probable errors vary between .03 and .05.

	MONROE (RATE)	MONROE (COMPREHENSION)	HAGGERTY SIGMA 3
Sangren-Woody	.73	.71	.70
Monroe (Rate)		.75	.56
Monroe (Comprehension)			.71

2. Tests of specialized reading skills are not valid tests of silent reading comprehension, as in the cases of the Briggs

English Form Test and of the Van Wagenen Reading Scales for English Literature.

TABLE 3

Validity coefficients for three silent reading tests

Note: The probable errors vary between .05 and .08. Numbers of cases are shown in parentheses below the coefficients.

	HAGGERTY	STANFORD
Van Wagenen Reading Scales for English Literature	.09 (90)	.34 (51)
Haggerty Sigma 3		.57 (71)

TABLE 4

Validity coefficients for six silent reading tests

Note: The probable errors vary between .02 and .05; the numbers of cases, from 66 to 70.

	SANGREN-WOODY	LOS ANGELES	ORLEANS	HAGGERTY	UNSTANDARDIZED	AVERAGE OF THE 5 STANDARD TESTS*
Stanford	.66	.70	.74	.86	.76	.85
Sangren-Woody		.69	.71	.74	.82	.85
Low Angeles Elementary			.66	.76	.71	.82
Orleans (Public School Achievement Test)				.75	.69	.74
Haggerty Sigma 3					.80	.89
Unstandardized†						.81

* The standard score unit was 1/10 S.D. Zero was taken as −3 S.D. −2.9 S.D. was given a score value of 1; −2.8, a score value of 2, etc. The standard score values of scores earned by any one student in all five standard tests were then averaged. In three or four instances, as few as three scores yielded by standardized tests were used. Absence from the school limited the number of tests taken by these pupils.

† This test was constructed by pupils in a class in educational measurements taught in the San Diego State Teachers College during the second semester of the school year of 1928–1929 by the first writer.

3. Carefully constructed home-made reading tests possess as good validity as standardized tests in the same subject matter field. This finding agrees with Ruch's comment that "the rank and file of such (standard) tests are readily equaled or bettered by the teacher who has mastered a little theory of measurement and who is seriously bent upon building valid and reliable examinations."[1]

[1] Ruch, G. M.: "The Objective or New-Type Examination." Chicago: Scott, Foresman and Company, 1929. Preface, iv.

WHO KNOWS ABOUT AVIATION?

VERNA KIMBERLEY WALLACE AND JOHN L. BROSE

State Teachers College, San Diego, California

If we should ask this question of "the man on the street," quite probably he would say that he knew something about flying and flyers. In an effort to discover whether such claims were just, we have carried out a short experiment involving adults, with one group of 83 students in the State Teachers College of San Diego representing the non-aviation "wing," and 45 students in two ground schools in San Diego, the technically trained group.

It is obvious that few other than technically trained persons would know about aircraft motors, aircraft systems, instruments used in controlling planes in flight, and so on. For this reason, the material used in the present study included only the nomenclature of aviation, outstanding events in the history of aviation, and air commerce regulations. The materials which were used in the study were obtained from the following sources: (1) "Nomenclature for Aeronautics," prepared by the National Advisory Committee for Aeronautics and published by the Government Printing Office at Washington, D. C.; (2) "Aviation—Fundamental Facts About Flying" (historical matter) prepared and distributed by the Haskin Information Bureau, Washington, D. C.; and (3) "Air Commerce Regulations," Instruction Paper 2 of the College of Engineering, Pacific Technical University, San Diego, California. The materials were presented to the subjects used in the experiment in the form of objective tests.

Excerpts from the three parts of the test follow.

Section 1. Nomenclature

Directions: If one of the following statements is true, place a check before the "T" on the left; if false, place a check before the "F". If in doubt, omit the item and go on at once to the next statement.

——T——F 2. Aspect foil means the ratio of span to mean chord of an air foil.

——T——F 2. The acute angle between the wing chords of a biplane or multiplane is termed decalage.

Etc.

Section 2. Outstanding Events

Directions: Read each characterizing phrase or date in the column at the right and then find the name of the man in the column at the left that the phrase best fits. Record the number of the correct event in the column at the right in the parenthesis in front of the correct name in the column at the left. You may find that one event is related to more than one man and that some men are not related to any of the events and dates.

( ) Byrd	1. First flight with passenger.
( ) Maitland	2. May 20–21, 1927.
( ) Bleriot	3. Round the world flight.

Etc

Section 2. Air Commerce

Directions: In each of the exercises below write in the word or words that have been omitted. Try to find the word or phrase for each blank that completes the statement correctly.

1. Air commerce means transportation in whole or in part by aircraft of ..... or ..... for hire.

4. ..... means a condition meeting the minimum aircraft requirements as set forth.

Etc.

The entire test included ninety opportunities to make correct scores. There were fifty exercises in section 1, twenty in section 2, and twenty in section 3.

The score distribution for the 83 untrained subjects ranged from 0 to 25 correct responses, with an average of 6.6 and a standard deviation of 4.2. This means that one-sixth (roughly) of the group had only two correct responses in the whole test, one-half had only six such responses and one-sixth had from eleven to twenty-five such responses. In the trained group the score range was from 0 to 35 with an average of 17.6 and a standard deviation of 9.1. One-sixth of this group had less than nine correct responses; one-half, seventeen or less; and one-sixth, twenty-seven or more.

An explanation of one cause of the apparently large number of low scores in the trained group is necessary. The test was given immediately following the Christmas vacation period (in January, 1930). In civilian aircraft training schools it is customary to organize the course into definite units, so that students can be accepted at frequent intervals, pursuing their work until they complete the cycle of units of instruction. There were some six or seven persons included in the trained group who had just entered the course, in one case having registered the day prior to the giving of the test. Instruction in the fields covered by the test is given during the first two weeks of study in all cases, however, so since the average number of weeks of study was 7½, with a range of 0 to 20, the test was fair to the majority of subjects.

Some further comments seem justified, concerning factors affecting scores. The form of the test used in presenting the materials was not familiar to the trained group but was familiar to the untrained group. The trained group took the test much less seriously than did the untrained group. Several persons in the trained group did not even complete the test, turning in their papers after only a few minutes of hurried work. Perhaps this was because papers were taken in without the students' names being required, so achievement on the test did not influence marks. On the other hand, the untrained group attempted every section of the test, and spent at least twenty-five minutes studying it before one paper was handed in.

The trained group clearly has an advantage over the untrained group. The difference between mean scores divided by the standard error of the difference is 7.7. Since a result of 3.0 is considered a reliable difference, we are assured that the result obtained, 7.7, is not one due to sampling, but is rather one such as one would always anticipate finding.

There are a number of interesting interpretations resulting from the experiment. One is that the average adult reads his newspaper and his informational periodicals rapidly, and without taking the trouble to look up unknown terms, even when the topic is as new and as interesting as is aviation. Another

is that the method of comparing trained and untrained groups is an excellent method for validating a test.[1] Still another is that when one enters a specialized activity, one should seek technical training in the various phases of that activity if one wishes to achieve proficiency.

[1] Ruch, G. M.: "The Objective or New-Type Examination." Chicago: Scott, Foresman and Company, 1929. Pages 27–28.

TEST PERFORMANCE OF A SELECTED GROUP OF PART-HAWAIIANS[1]

C. M. LOUTTIT[2]

Psychological Clinic, University of Hawaii

The Psychological Clinic of the University of Hawaii is engaged on a study of the mental ability of Hawaiian hybrids, with financial aid from the Rockefeller Fund for the Study of Race. As part of this larger program the writer gave the Binet, Porteus Maze and Healy Picture Completion No. II tests to all of the students in the Kamehameha Schools at Honolulu. This paper presents the results of this testing program showing the test performance by age, grade and race.

The Kamehameha Schools include a School for Boys, a School for Girls, and a Preparatory Department (from which no students were tested). At present the policy is to admit only students with some degree of Hawaiian blood. All entering students are graded according to the results of the Otis Classification test, the National Group Intelligence test and their previous academic record. Unlike many private schools there is no selection on the basis of finances. In fact, other things being equal, the student in the poorer financial circumstances is given preference. From the point of view of the administration the selecting factors in both schools are identical, but, as will be shown later, the students of the girls' school are slightly below those of the boys' school in test performance. Thus the

[1] For the opportunity to use these students I am indebted to Mr. Frank E. Midkiff, President of the Schools, and for assistance in carrying out the work I wish to thank Mr. John F. Nelson, Principal of the Boys' School, Miss Maude Schaeffer, Principal of the Girls' School, and Mr. George A. Odgers, Director of Tests and Measurements.

[2] Now at Ohio University.

subjects used in this research are a selected group of secondary school pupils all of whom have some Hawaiian blood.[3]

The Porteus Maze (hereafter referred to as the Maze test) and the Healy Picture Completion No. II (hereafter referred to as the Healy) were given according to the standard published directions. The Binet, however, was the modified form of the Vineland revision described by Porteus and Babcock (1926). In brief this modification is "designed to eliminate as far as possible those tests immediately dependent upon language and those that require an oral response in well phrased English" (Babcock, 1927).

AGE

It has been the regular practice of the Psychological Clinic to use 14 years as the upper limit of mental development when calculating the IQ for persons above that age. The present data for various ages, as shown in table 1, affords support for this practice. There is no uniform increase in the ability on the Binet above 14 years for either boys or girls, in spite of the fact that this is a selected group. A similar circumstance holds for the girls in the case of the Maze, while the boys' average in this test remains fairly constant around 15 years' performance.

Except for ages 16 and 17 in the Binet the boys excel the girls to a greater or less extent in all tests. The difference, however, is hardly to be ascribed to a real sex difference, but is probably largely due to a factor of selection not under control of the school. It is impossible to measure this factor or even to suggest what it might be.

It is perhaps of interest to compare the performance of this group with that of an unselected one. While there are no data on a strictly similar group in point of age, Porteus and Babcock (1926) have published the following scores for unselected 14 year-old Part-Hawaiian boys: Binet, 35 cases, average mental

[3] The only exceptions are one boy who is pure Samoan and one girl who is Samoan-white. These two cases are included in the data presented.

age 11.25, average IQ 80.0; Maze, 33 cases, average test age 13.06, average M.Q. 93.3. Furthermore the average IQ for 9, 12, and 14 year-old Part-Hawaiian boys on the Binet was 84.33;

TABLE 1

Average test scores for each age

CHRONOLOGICAL AGE	NO. OF CASES	BINET				MAZE				HEALY	
		Age	σ	I.Q.	σ	Age	σ	M.Q.	σ	SCORE	σ
Boys											
13	4	14.91	.66	109.00	4.06	14.75	.90	108.00	2.00	83.50	3.18
14	11	14.05	1.32	100.45	9.19	14.95	1.15	106.81	7.79	75.77	11.13
15	39	13.98	1.55	99.79	8.53	15.05	.93	107.51	6.39	78.18	12.62
16	36	14.05	1.28	100.58	8.73	15.18	.87	108.39	5.99	76.55	10.13
17	37	13.78	1.27	98.51	8.85	14.71	1.38	105.16	9.85	74.09	16.18
18	32	14.67	1.05	104.62	7.29	15.33	.83	109.15	5.54	70.85	14.84†
19	28	14.13	1.27	100.93	8.72	15.14	1.04	108.10	6.96	75.47	14.53‡
20	16	14.45	1.26	103.25	8.72	14.81	1.29	105.69	9.15	68.70	16.28§
21*	21	13.76	1.79	98.81	12.39	15.52	.59	110.76	4.07		
Girls											
11	3	12.21		105.33		12.66		108.66		43.66	
12	9	12.23	1.00	97.44	7.67	14.55	1.44	116.11	11.65	58.39	18.05
13	13	12.92	1.37	94.84	10.17	13.54**	1.59	98.66	11.54	57.27	17.79
14	22	13.08	1.55	93.54	10.80	14.27	1.69	101.95	11.98	66.54	13.86
15	25	13.80	1.24	98.56	8.62	14.48	1.21	103.48	8.52	67.46	15.70
16	23	14.18	.99	101.43	6.82	14.61	1.19	104.43	8.22	75.72	13.00
17	23	14.27	1.31	102.13	9.07	14.28	1.31	102.21	9.12	63.74	14.84
18	15	13.68	1.27	97.86	9.00	14.30	1.16	102.33	8.06	63.20	24.04
19	2	13.92		100.00		12.0		85.50		72.50	
20	1	14.10		101.00		14.00		100.00		75.00	
21	1	11.60		83.00		16.00		114.00		49.50	

* Average age 21.52 ± .79.

† 27 cases.

‡ 18 cases.

§ 10 cases.

** Only 12 cases.

the average M.Q. was 100.14. These data from public school boys are evidence that the Kamehameha students are a selected group.

On the Healy both the boys and girls did as well or better than the norms published by Healy (1921) with the exception of the girls in three age groups, only one of which had an adequate number of cases to make a comparison.

GRADE

Table 2 presents the averages for the three tests by grade. Both boys and girls show an increase of several points in the Binet IQ from the 7th to the 12th grade. The girls in the lower grade are, however, superior to those in the higher in Maze performance. This is not true for the boys who show a slight increase in the M.Q.[4] Performance on the Healy shows improvement in the case of the girls to the 11th grade after which there is a decided lowering in the score. There is no consistent change in the boys' performance from the 7th to H-11th grade. It is interesting to note the difference in age between the boys and girls. The latter range from an average of 12.5 years in the 7th grade to 17.4 in the 12th, while the boys' averages for the same grades are 14.7 and 20.5 years.

RACE

For the purpose of comparing the various races the scores of the boys and girls were averaged together and the data of table 4 based on these total averages. The values for each group are given in table 3. There were only four hybrid groups with sufficient number of cases to warrant statistical treatment, viz., (1) Hawaiian (*H*)—individuals who claimed pure Hawaiian ancestry were included, although it is impossible to be certain that there is no admixture of foreign blood. (2) Hawaiian-white

[4] It must be noted that the H-11 (High 11th) and 12th grades in the boys' school are really one grade. The boys spend half time in school and half time working in and around Honolulu. Thus two years are required to finish the last year.

TABLE 2

Average test scores for each grade

GRADE	NUMBER OF CASE	BINET				MAXE				HEALY		CHRONOLOGICAL AGE	
		Age	σ	I.Q.	σ	Age	σ	M.Q.	σ	Score	σ	Years	σ
Boys													
7	21	13.65	1.39	98.00	10.26	15.02	1.05	107.76	7.12	76.71	12.47	14.71	.88
8	24	13.44	1.31	96.17	8.41	14.37	1.01	105.79	6.47	74.75	13.20	15.58	1.03
9	49	13.95	1.26	99.75	8.76	14.91	1.22	106.36	8.63	75.60	12.34	16.12	1.08
10	39	14.62	1.48	104.36	8.94	15.28	.79	109.08	5.57	77.02	14.06	16.74	1.17
Low 11	31	14.47	1.15	103.32	7.99	14.88	1.19	106.29	8.42	76.32	13.79	18.09	1.11
High 11	34	14.33	1.39	102.26	9.73	15.42	.78	110.02	5.39	78.77*	15.24	19.38	1.33
12	26	14.39	1.43	103.11	9.81	15.29	1.04	109.08	7.25			20.57	1.24
Girls													
7	13	11.95	.93	93.69	10.15	13.57	1.64	106.15	15.06	53.23	18.21	12.54	1.15
8	17	12.78	1.39	94.94	11.33	14.12†	1.72	105.00	13.04	58.41	14.76	13.18	.98
9	24	12.97	1.19	93.21	8.73	14.06	1.44	100.83	10.52	70.64	12.92	14.54	1.00
10	27	13.79	1.03	98.77	7.04	14.42	1.26	103.22	8.71	69.24	14.40	15.52	.99
11	29	14.48	1.09	103.55	7.79	14.55	1.11	104.20	7.84	69.94	19.27	16.03	1.40
12	27	14.22	1.09	101.69	8.66	14.35	1.46	102.59	10.21	62.42	16.03	17.44	1.75

* Only 18 cases.

† Only 16 cases.

(*HW*)—any percentage of Hawaiian with North European. (3) Hawaiian Chinese (*HC*)—Any percentage of Hawaiian with Chinese. (4) Hawaiian-white-Chinese (*HWC*)—any percentage of Hawaiian with North European and Chinese. All other mixtures were omitted from the tables of this section but were included in all previous tables. The other mixtures include Japanese, Portugese, Negro, Samoan, Korean, and American Indian with Hawaiian.

The *HWC* boys took first place in the three tests; the girls first in the Binet and Maze, but fall to third in the Healy. The pure Hawaiian girls are last in all tests, while the boys are second in the Maze, third in the Binet, and last in the Healy. *HW* girls are second in all tests, and the boys are second in Binet and Healy but third in the Maze. The *HC* girls are first in the Healy and last in the Binet and Maze. Arranged in descending order of average rank for boys and girls the race mixtures stand; *HWC*, *HW*, *HC*, *H*.

The variability of the groups may be summarized from table 3 as follows. *H* boys are most variable in the Binet and Healy and are third in the Maze, while the girls are first in the Healy, second in the Maze and third in the Binet. The *HW* boys are second in the Maze, third in the Binet and last in the Healy; the girls are most variable in Binet and Maze and second in the Healy. *HC* girls are least variable in all three tests, but the boys are most variable in the Maze and second in the Binet and Healy. The *HWC* boys are least variable in the Binet and Maze and third in the Healy, while the girls are second in the Binet and third in the Maze and Healy. The race mixtures arranged in order of decreasing variability are *H*, *HW*, *HC*, *HWC*.

It is also of interest to note that the variability on the Maze and Binet is of approximately the same order of magnitude, while that of the Healy is from two to three times greater.

In table 4 are presented the data relative to the significance of the differences between the averages of the several race groups. This table is read as follows. In the second row of the third column we find the difference between *HC* and *HW*

TABLE 3

Average test scores for each race mixture

RACE	NUMBER OF CASES	BINET				MAZE				HEALY			CHRONOLOGICAL AGE	
		Age	σ	I.Q.	σ	Age	σ	M.Q.	σ	Cases	Score	σ	Years	σ
Boys														
H	21	13.63	1.41	97.52	9.81	15.11	.89	107.95	6.19	19	68.31	15.17	17.28	2.35
HW	104	14.38	1.22	102.73	8.47	15.09	1.02	107.81	7.09	80	77.36	12.08	17.36	2.22
HC	48	13.55	1.33	96.91	9.52	14.85	1.37	106.10	10.46	41	76.32	13.88	17.12	2.01
HWC	17	14.62	1.04	104.70	7.19	15.29	.79	109.41	5.32	13	78.88	12.62	15.35	3.30
Girls														
H	7	12.03	.43	88.60	4.10	13.10	.82	95.90	9.35		51.00	27.00	15.00	2.20
HW	70	13.61	1.39	98.30	9.39	14.40*	1.46	104.20	11.68		66.05	16.98	15.24	1.72
HC	25	13.60	1.32	98.00	3.16	14.26	.82	102.80	5.92		68.40	13.43	15.64	1.78
HWC	12	13.80	1.36	101.30	7.44	14.60	1.08	106.60	10.30		60.00	14.25	15.50	2.32

* Only 69 cases.

TABLE 4

Comparisons between the various race mixtures

RACE	NUMBER OF CASE	AVERAGE I.Q.	σ	H	HW			HC			HWC		
					x	σ	$\frac{x}{\sigma}$	x	σ	$\frac{x}{\sigma}$	x	σ	$\frac{x}{\sigma}$
H	28	95.25	8.74		5.70	1.78	3.20	2.03	1.89	1.07	8.04	2.13	3.77
HW	174	100.95	8.85	9993				−3.63	1.15	3.15	2.34	1.51	1.55
HC	73	97.28	7.94	8517		9991					6.01	1.64	3.66
HWC	29	103.29	7.29	9999		9394			9998				
Maze													
H	28	104.79	7.11		1.57	1.51	1.04	.18	1.71	.105	3.45	1.96	1.76
HW	173	106.36	9.28	8508				−1.39	1.28	1.09	1.88	1.60	1.17
HC	73	104.97	9.16	5438		8621					3.27	1.79	1.82
HWC	29	108.24	7.78	9608		8790			9656				
Healy													
H	26	63.65	19.11		8.43	3.92	2.15	9.67	4.11	2.35	6.16	4.61	1.33
HW	150	72.08	14.41	9842				1.24	2.04	.608	−2.27	2.93	.775
HC	66	73.32	13.71	9906		7291					−3.51	3.17	1.10
HWC	25	69.81	13.44	9082		7794			8643				

If the rows exceed the column the difference is minus (−).

is 3.63 in favor of the *HW*, the $\sigma_{diff.}$ is 1.15 and the $\frac{x}{\sigma}$ value is 3.15. The value 9991, in the third row of the second column represents the chances in 10,000 that the difference is greater than 0.

In the Binet the chances of the differences being significant are all better than 99 in 100, except that between *HWC-HW* where it drops to 93, and between *HC-H* where it is only 85. In the Maze none of the differences give a probability of 99 in 100; those between *HWC-H* and *HWC-HC* being only 96, between *HC-H* only 54, while the others are between 85 and 87. In the Healy only the difference between *HC-H* gives a probability of 99 in 100; *HW-H* 98 and *HWC-H* 90. The *HC-HWC* difference drops to 86 while the others are 72 and 77.

Considering the combined averages as shown in this table we find the *HWC* leading in the Binet and Maze while the *HC* take first place in the Healy. The *HW* take second place in the three tests. Third place is taken by the *HC* in the Binet and Maze and by the *HWC* in the Healy. Again the *H* occupy last place in all three tests. These ranks satisfactorily check those given a few paragraphs above. A further evidence of *HWC* superiority is to be noted in the fact that this group is approximately one year younger than any of the other groups.

SUMMARY

1. Three tests—Binet, Porteus Maze, and Healy Picture Completion No. II—were given to 224 boys and 137 girls with Hawaiian blood at the Kamehameha Schools, Honolulu.

2. The average performance of all the tests remains fairly constant at 14 years giving some evidence that this level must be considered the upper limit of mental development—at least for those abilities measured by the tests.

3. The higher grades show superior ability in the Binet for both boys and girls and in the Healy for the boys. The boys on the Maze and Healy show no consistent increase in ability. On the Maze the girls are superior in the lower grades.

4. *a.* The differences in Binet performance for the different race groups are all significant except between the *HC* and *H*.

b. For the Maze scores only the difference between *HWC-H* and *HC-HW* are fairly significant.

c. The comparison of the *H* with all other groups is fairly significant in the Healy, but the differences for this test are not significant for any other groups.

REFERENCES

BABCOCK, M. E. 1927 Application of clinical psychology in Hawaii. Univer. of Haw. Res. Pub. No. 1,

HEALY, WM. 1921 Picture Completion test II. Jour. Appl. Psychol., 5:225–239.

PORTEUS, S. D., AND BABCOCK, M. E. 1926 Temperament and race. Boston: Badger.

A COMPARISON OF WHITE AND NEGRO CHILDREN IN RHYTHM AND CONSONANCE[1]

ROSALIND STREEP

Hunter College

I. INTRODUCTION

With regard to the musical ability of the negro as compared with that of the white, an *a priori* assumption has been and is to a certain extent now held. It has been more or less generally felt, without the support of experimental evidence, that the negro is superior to the white in musical ability. This opinion apparently has been based on the everyday manifestations of the two races in this particular phase of mental life. However, for purposes of scientific value the truth or falsity of this assumption is to be found only in the facts of experimental research.

In view of the complexity of the two races the investigator is forced to concern himself not only with the tangible factors of large numbers and the statistical computations thereof, but also with the influence of the highly intangible factors of heredity and environment. Not until such factors as the educational and social advantages of the two races have been determined

[1] This study was undertaken as an M.A. thesis at Columbia University.

The writer wishes to acknowledge her indebtedness: To Professors M. T. Whitley of Teachers College, A. T. Poffenberger and H. E. Garrett of Columbia University for their helpful suggestions, constant readiness to aid, and many kindnesses; to Dr. William J. O'Shea, Superintendent of New York City Public Schools, to Mr. Eugene A. Nifenecker, Director of the Bureau of Reference, Research and Statistics, New York City Board of Education, to the various principals in whose schools this investigation was conducted, and to the teachers whose grades were tested, for their coöperation in this study.

will a comparison of these races be reliable. A statement by Garth (2) may serve to confirm the above: "The elements in the study of racial mental similarities or differences must be these: (1) Two so-called races, R_1 and R_2; (2) an equal amount of education and opportunity, E, which should include social pressure and racial patterns of thought; and (3) psychological tests, D, within the grasp of both racial groups. We should have as a result of such an experiment R_1ED equal to, greater than, or less than R_2ED."

Concerning the influence of social heredity and environment upon the negro, Peterson and Lanier (5) indicate their views in the following: "These negroes have had a unique social and economic history which may wholly account for their characteristic features in musical behavior. The effects of different musical backgrounds on responses to mechanically produced stimuli like the Seashore tests are not yet known. It is certain that the results in the consonance test are not independent of the subjects' cultural background." In addition, to quote Peterson (7): "Consonance preferences are considerably influenced, we find, by numerous overlapping attitudes and sets engendered by the general relationships of tones under which the intervals reacted to are presented."

In view of the fact that it is so difficult to regulate and control these factors, results attained concerning racial differences will probably be subject for the most part to a discussion of the relative influence of heredity and environmental factors, with chief emphasis on a consideration of biological, cultural, and social factors.

Since the assumption concerning the superiority of negroes over whites has been based chiefly on manifestations of the sense of rhythm and that of consonance, the present report confines itself to an investigation of these phases of musical ability. However, it must be borne in mind that the abilities tested—"senses," as Seashore terms the phases of the musical mind (9)—are regarded as specific capacities, each in a more or less peculiar way contributing to success in musical art.

II. REVIEW OF WORK PREVIOUSLY DONE

In order to give a proper setting for any study it is well to review briefly some of the results that have been attained in previous studies concerning the same subject. A comparatively large amount of work concerning differences between the white and negro has been done in relation to intelligence, learning, resistance to fatigue, and personality. However, in relation to the comparative abilities of whites and negroes in music or certain phases of music, very little by way of experimental research has been done.

Professors Peterson and Lanier conducted an investigation concerning the comparisons of certain mental abilities in white and negro adults (5). Among other tests of mental ability, the Seashore Measures of Musical Talent were administered to the two groups. "The white group tested were students of the Middle State Teachers College at Murfreesboro, and the negro group were students of the Agricultural and Industrial Normal College, Nashville. . . . . We have included in our study only subjects from the last two years of high school and the college classes." It may be well to represent here the results of this investigation in tabulated form as shown in the monograph.

Comparisons of median scores in per cent right by white and negro adults (negative differences favor negroes)

TESTS	WHITES		NEGROES		DIFFERENCE	PER CENT OF WHITES ABOVE MEDIAN OF NEGROES
	Number	Median	Number	Median		
Pitch	372	64.01	288	57.93	6.08	72.3
Intensity	373	84.70	296	75.60	9.10	80.5
Time	372	76.59	295	70.83	5.76	75.5
Consonance	386	66.70	288	63.40	3.30	59.5
Tonal Memory	382	55.43	289	43.36	12.07	78.5
Rhythm	385	70.13	290	70.21	−.08	49.7

The results, as indicated by the table, tend to show a superiority of the white adult over the negro adult in all phases of musical ability except rhythm. In this last-mentioned phase

the difference in favor of the negroes is so small that it may be considered statistically unreliable.

Aside from the investigation described above no additional studies concerning the comparative musical abilities of whites and negroes have been found within the scope of published psychological material.

III. PURPOSE OF THE INVESTIGATION

The purpose of the present investigation is three-fold. An attempt is made, first, to determine whether differences in the senses of rhythm and consonance between white and negro children exist, and, if so, the extent of these differences; second, to determine whether the sense of rhythm bears any relation to intelligence, and, likewise, whether the sense of consonance bears any relation to intelligence, and, if so, the extent of the relationship; and third, to determine whether the application of the Seashore tests of rhythm and consonance is suited to children of the two elementary school grades below the fifth; i.e., children of the third and fourth grades.

IV. THE SUBJECTS TESTED

The present report is based on tests of 1315 children of the New York City public schools, of which 637 were white children, and 678 were negro children. The white group tested were pupils of Public School 165 (Manhattan), and the negro group were pupils of Public Schools 5 (Manhattan) and 157 (Manhattan).

The choice of subjects for this investigation was determined by the desire to select as comparable groups of the two races as possible. The method followed was to select schools which were considered representative or capable of yielding "fair samplings." In accordance with this plan and also with the purpose of the investigation, children of the third, fourth, fifth, and sixth grades of the respective schools were tested. The grades of the New York City Public School System are classified on the same basis; i.e., first, grades 3, 4, 5, and 6 include, in all schools, approximately, children of chronological ages

eight through twelve; and, second, every grade in the system is divided into three types of groups: "bright normal" (I.Q. 110+), "normal normal" (I.Q. 109–90), and "dull normal" (I.Q. 89−). In view of this classification the selection made for this study appeared to be the most representative and consequently capable of yielding the most reliable results.

Table 1 shows the number of children for each grade classified according to race.

Not only were the two tests administered to equal numbers within each grade as indicated by table 1 but to those children to whom the rhythm test was administered, the consonance test was also given.

The basis for the selection of negro and white children was skin color. The presence of negro blood was judged by ex-

TABLE 1

Classification of subjects taking the tests

SEASHORE MUSIC TESTS	THIRD GRADE	FOURTH GRADE	FIFTH GRADE	SIXTH GRADE	TOTAL
Rhythm and consonance:					
White	166	157	160	154	637
Negro	184	186	136	172	678

ternal evidence; i.e., skin color served as the biological differential upon which the selection was made. Colored hybrids and negroes apparently approaching pure-blood stock were classified as a whole as "negroes" for the purpose of this investigation.

V. THE TESTS AND THEIR ADMINISTRATION

The tests used were the rhythm and consonance tests of the Seashore Measures of Musical Talent. Phonograph records which record the tests were played on a portable victrola. The testing was conducted during the morning school hours of May and June, 1929. All the testing was done in rooms of the schools from which the children were selected. In all cases the rooms were well lighted and ventilated and as quiet as possible. However, the noises that did occur were beyond the immediate

control of the school officers; as, for example, those resulting from street traffic.

The writer of this report, a white person, administered the tests to groups of about forty in number. Teachers of the respective classes lent their coöperation by seeing to it that the directions were followed throughout the test. Uniformity in the administration of the tests was assured by the fact that the same person administered the tests to every group of children.

On each of the two tests, fifty judgments were recorded by each subject. The judgments were recorded on standard printed blanks, each containing 100 squares (10 columns and 10 rows (8). The score on each test is the per cent of correct judgments.

Preparatory to the playing of the record for a given test, the experimenter illustrated by means of a practice exercise the way in which judgments were to be made. The attitude of both experimenter and subjects toward the tests was not one of "tests" but of "games." The subjects were asked to call out aloud in unison the judgments for, in the test of rhythm, the first four trials; and in the test of consonance, the first three trials. The purpose of this practice exercise was to familiarize the subjects with the type of discrimination to be made and to prevent as far as possible a misunderstanding of the directions. After the practice exercise for a given test, the experimenter stated the directions given by Seashore for the test (8). The time for the administration of each test including preliminary exercises was about twenty minutes.

The tests were given in two sittings. The rhythm test was given at the first sitting, and on the following day the consonance test was given to the same pupils. Both tests were never given on the same day.

Intelligence test scores, which were used for the purpose of correlation ratings, were obtained from school records.

VI. DATA

The comparison between the two races is based upon both a separate comparison of white and negro children of correspond-

ing grades and a comparison of white and negro children as a whole.

1. Comparison of white and negro children of corresponding grades

In regard to this phase of the study, the bases for comparison are the median, the standard deviation, the amount of overlapping in the distribution of scores made by the two groups, and the coefficient of variation. The latter measure was computed by means of Pearson's Coefficient of Variation, given by the formula,

$$V = \frac{100\,\sigma}{\text{Average}}$$

In addition, the correlations between intelligence and the two Seashore tests are based upon the separate grade classification. The coefficients of correlation were computed by means of the Spearman Rank Method, expressed by the formula,

$$r = 1 - \frac{6\,\Sigma\,D^2}{N(N^2 - 1)}$$

The results obtained concerning the grade comparison in rhythm and in consonance are given in tables 2 and 3, respectively. Accompanying charts serve to a certain extent as graphic representations of the tabulated results.

2. Comparison of white and negro children as a whole

In regard to this phase of the study the bases for comparison are the respective averages of the four medians found for each race, and the respective averages of the four 75th percentile scores found for each race. The results obtained concerning the total comparison of the two races in rhythm and consonance are given in table 4.

VII. RESULTS

As indicated by table 2 the median score made by negro children on the rhythm test in each grade exceeds to a small

TABLE 2

Grade comparison of white and negro children in rhythm

SCHOOL GRADE	NUMBER		MEDIAN		DIFFERENCE IN FAVOR OF NEGROES	STANDARD DEVIATION		PER CENT OF ONE RACE ABOVE MEDIAN OF THE OTHER		PER CENT OF ONE RACE ABOVE 75TH PERCENTILE OF OTHER		COEFFICIENT OF VARIATION		COEFFICIENT OF VARIATION (NEGRO) DIVIDED BY COEFFICIENT OF VARIATION (WHITE)	CORRELATION WITH I.Q.	
	White	Negro	White	Negro		White	Negro	White above negro	Negro above white	White above negro	Negro above white	White	Negro		White	Negro
3	166	184	62.50	62.90	.40	7.80	8.20	47	53	17	33	12.48	11.45	.92	+.13	+.24
4	157	186	64.79	67.58	2.79	8.90	7.55	40	60	14	36	13.74	11.17	.81	+.16	+.15
5	160	136	70.51	71.38	.87	7.20	7.70	45	55	19	31	10.21	10.78	1.05	+.10	+.10
6	154	172	72.30	73.20	.90	8.05	7.20	45	55	19	31	11.13	9.84	.88	+.05	+.11

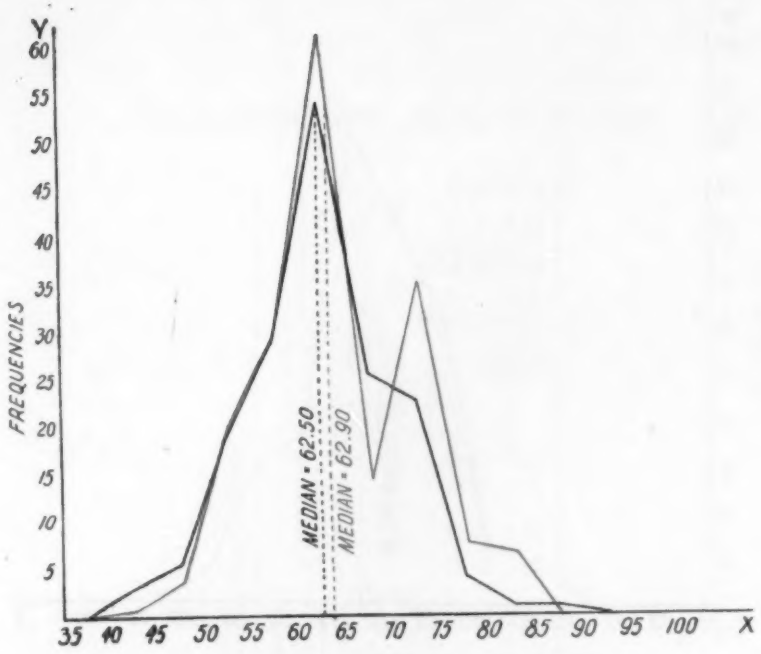

FIG. 1. RHYTHM SCORES, THIRD GRADE
Black: White children. Red: Negro children

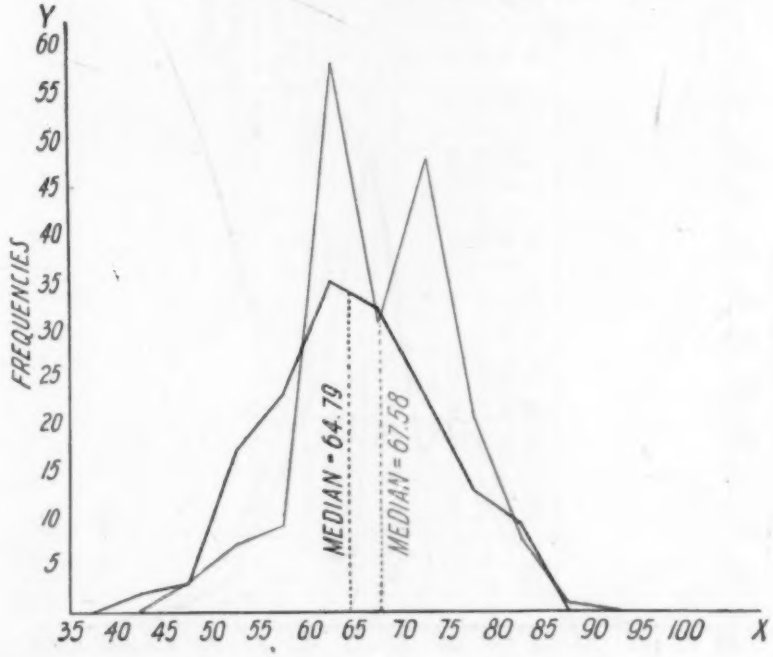

FIG. 2. RHYTHM SCORES, FOURTH GRADE
Black: White children. Red: Negro children

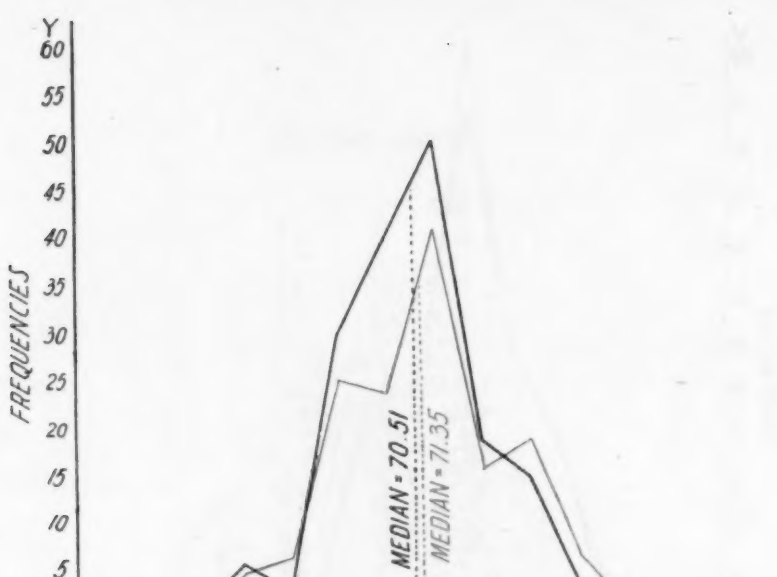

FIG. 3. RHYTHM SCORES, FIFTH GRADE
Black: White children. Red: Negro children

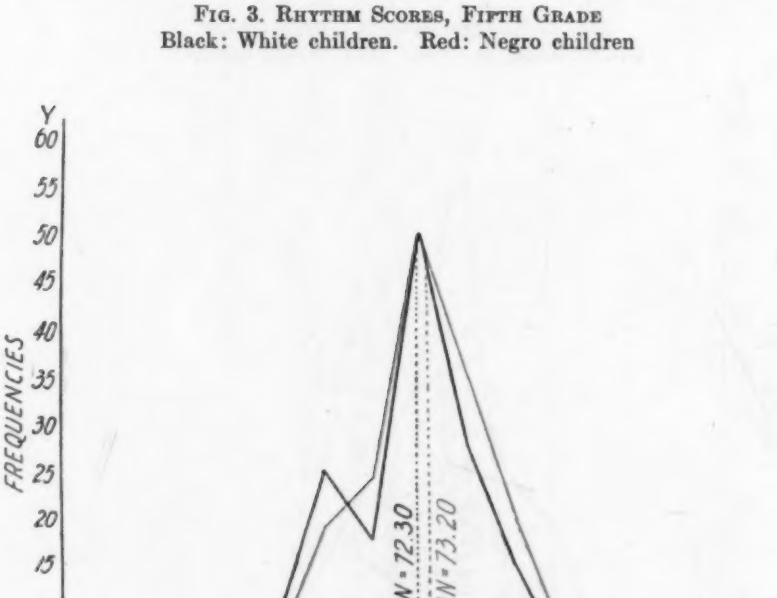

FIG. 4. RHYTHM SCORES, SIXTH GRADE
Black: White children. Red: Negro children

TABLE 3

Grade comparison of white and negro children in consonance

SCHOOL GRADE	NUMBER		MEDIAN		DIFFERENCE IN FAVOR OF NEGROES	STANDARD DEVIATION		PER CENT OF ONE RACE ABOVE MEDIAN OF THE OTHER		PER CENT OF ONE RACE ABOVE 75TH PERCENTILE OF OTHER		COEFFICIENT OF VARIATION		COEFFICIENT OF VARIATION (NEGRO) DIVIDED BY COEFFICIENT OF VARIATION (WHITE)	CORRELATION WITH I.Q.	
	White	Negro	White	Negro		White	Negro	White above negro	Negro above white	White above negro	Negro above white	White	Negro		White	Negro
3	166	184	58.41	60.89	2.48	8.30	8.25	40	60	17	33	14.21	13.55	.95	+.06	+.10
4	157	186	60.94	62.20	1.26	8.55	7.90	43	57	18	32	14.03	12.70	.91	0	+.17
5	160	136	63.30	64.49	1.19	7.15	7.45	44	56	19	31	11.29	11.15	.98	+.13	−.13
6	154	172	64.58	65.38	.80	10.01	7.15	46	54	27	23	15.50	10.94	.71	+.20	+.25

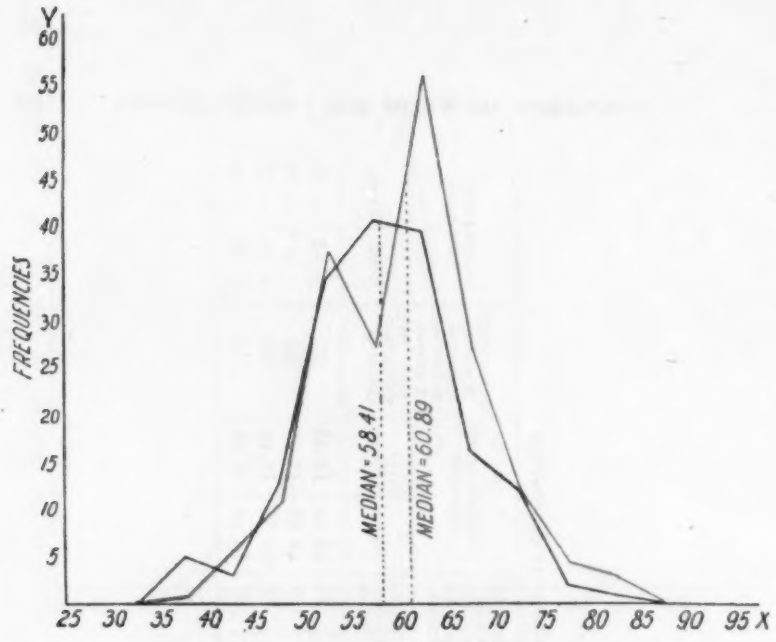

FIG. 5. CONSONANCE SCORES, THIRD GRADE
Black: White children. Red: Negro children

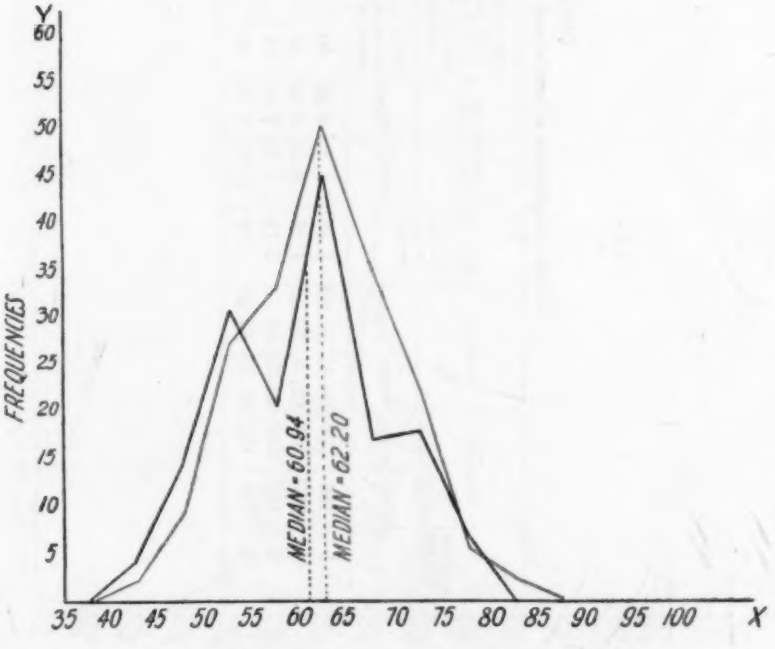

FIG. 6. CONSONANCE SCORES, FOURTH GRADE
Black: White children. Red: Negro children

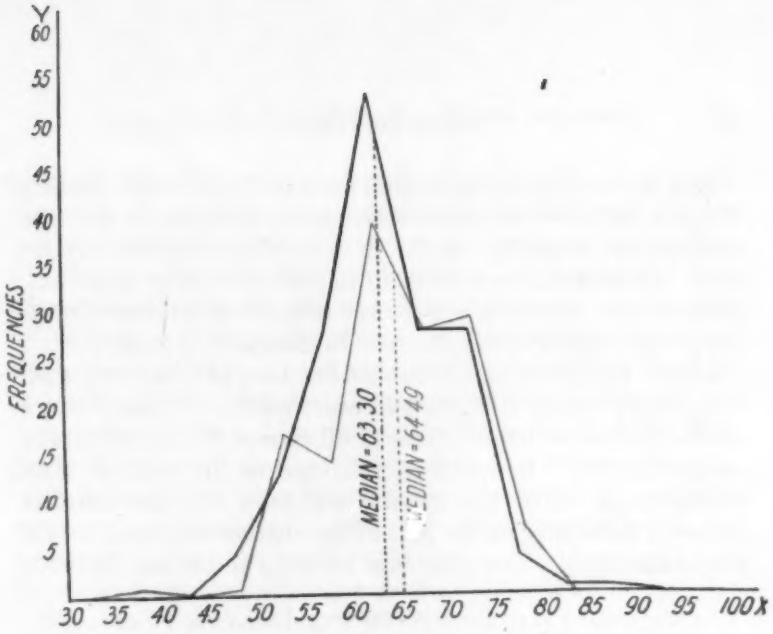

FIG. 7. CONSONANCE SCORES, FIFTH GRADE
Black: White children. Red: Negro children

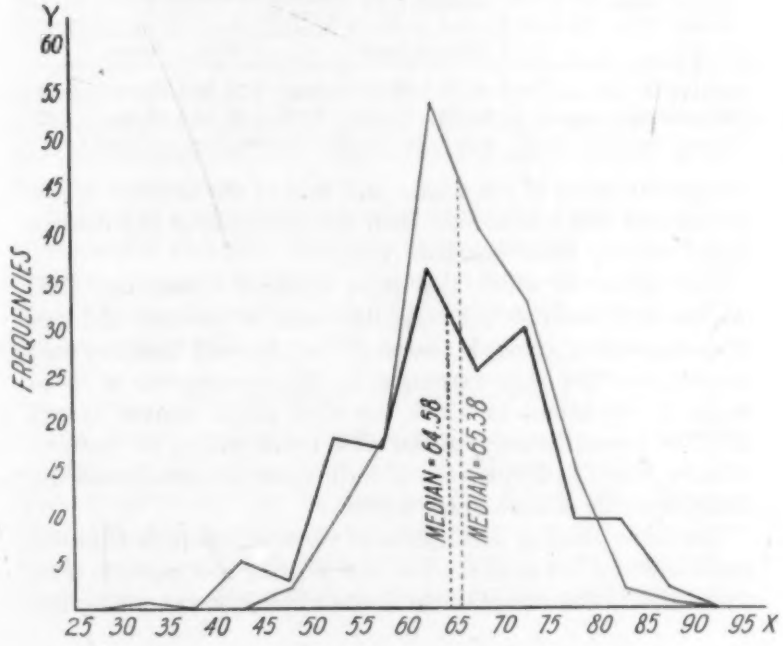

FIG. 8. CONSONANCE SCORES, SIXTH GRADE
Black: White children. Red: Negro children

degree the corresponding median score made by white children. Table 3 indicates results concerning performance on the consonance test markedly similar to those yielded by the rhythm test. These results would seem to indicate a very slight but, nevertheless, consistent superiority of negro children over white children in regard to the phases of musical ability tested.

It is of interest to note here that Peterson and Lanier in their comparative study of white and negro adults (5) found negro adults inferior to white adults in all phases of musical ability except rhythm. In addition, reference to the table of their results (page 55 of this report) will show that the smallest degree of inferiority on the part of the negroes was found in the consonance test. The difference favoring the whites between

TABLE 4

Total comparison of white and negro children in rhythm and consonance

TESTS	AVERAGE OF MEDIANS OF GRADES 3, 4, 5, 6		DIFFERENCE IN FAVOR OF NEGROES	AVERAGE OF 75th PERCENTILE SCORES OF GRADES 3, 4, 5, 6		DIFFERENCE IN FAVOR OF NEGROES
	White	Negro		White	Negro	
Rhythm	67.53	68.77	1.24	75.03	77.04	2.01
Consonance	61.53	63.24	1.71	67.55	68.48	.93

the median score of the whites and that of the negroes on the consonance test was smaller than the difference of the median scores on any other Seashore test.

The extent to which the negro children exceed both the median score and the 75th percentile score of the white children of corresponding grades is also small but, in every case but one, consistent. The only exception to this consistency is to be found in the extent to which the sixth grade negroes exceed the 75th percentile score of the sixth grade whites on the consonance test; i.e., 23 per cent of sixth grade negroes exceed the 75th percentile of sixth grade whites.

The corresponding coefficients of variation indicate that the negro children are in all but one case slightly less variable than the whites. The exception to this is to be found in the results

obtained for fifth grade negroes in rhythm. The fifth grade negroes are 5 per cent more variable than the fifth grade whites.

The results obtained concerning the correlation between intelligence and rhythm and that between intelligence and consonance indicate in all but two situations a positive but low relationship. The two exceptions to the positive correlation are to be found first, in the zero correlation between the I.Q. of fourth grade whites and their performance in consonance; and, second, in the $-.13$ correlation between the I.Q. of the fifth grade negroes and their performance in the consonance test.

As indicated by table 4 the results concerning the comparison of negro and white children as a whole serve to verify the results concerning the grade comparisons. The negro children on the whole are but slightly superior to the white.

The actual difference between the medians of each grade and for each test points in favor of the negroes. However, whether this difference is sufficiently large to insure a true difference greater than zero in favor of the negroes must be considered. The statistical computation of this fact serves to show that:

1. For the sense of rhythm the chances that there is a difference greater than zero between the true median scores of the respective grades are 64 in 100 for the third grade, 99.2 in 100 for the fourth grade, 79 in 100 for the fifth grade, and 80 in 100 for the sixth grade.

2. For the consonance test the chances that there is a difference greater than zero between the true median scores of the respective grades are 98.7 in 100 for the third grade, 87 in 100 for the fourth grade, 86 in 100 for the fifth grade, and 74 in 100 for the sixth grade.

VIII. SUMMARY AND CONCLUSIONS

The foregoing pages present in comparison the results achieved by 637 white and 678 negro children working at the same problem each in his own familiar school environment. The subjects concerned in this test were not the races as a whole but only those representatives serving as students in the New York City elementary schools under the conditions men-

tioned. However, since both the negro and white samplings are from New York City, the two groups may be considered comparable if not truly representative of the two races. Thus, it may be seen that an attempt to control the environmental factor was made in that the investigation represents results of the comparison of white and negro children in the same section of the country and in the same school system.

The results obtained indicate a slight superiority of the negroes over the whites in the two phases of musical ability tested. The results indicate, furthermore, in view of the degree of overlapping in the distributions of scores by the two races, a greater variability within each race than between the races; i.e., the individual differences within either group—between negro and negro or between white and white—are probably of more importance, because they are greater, than the differences between negro and white. That the variability of the scores of these children may be due to race differences is highly doubtful. The variation is probably due to variations in individual capacities.

Race differences that are in evidence as far as the results of these tests are concerned are negligible for practical purposes. However, it is of particular interest to note that although the superiority of the negro in these two phases of musical ability is small, we find, although not equivalent numerically, a corresponding superiority throughout the respective grades. Such differences as are obtained may be due to the influence of different social environments demanding different attitudes and activities.

It may be said that if one were to test the musical capacity of the unselected masses of negroes, with their smaller percentage of white blood, and make a comparison with the unselected masses of whites, the results might possibly reveal greater racial differences than those found herein.

The results obtained concerning the correlation of rhythm and consonance with intelligence serve to substantiate Seashore's contention (9) that ability in any one of the musical "senses" cannot be regarded as an index to general intelli-

gence but, rather, as an indication that the specific musical sense requires a certain degree of general intelligence. The two exceptions, mentioned on page 67, suggest statistical unreliability in that fourteen out of sixteen measures of correlation are positive.

In addition, in view of the median scores of third and fourth grade children on both tests, it may be said that the application of the Seashore tests of rhythm and consonance is suited to children of at least two elementary school grades below the fifth.

In addition to the foregoing conclusions a deviation from the general tendency of the results demands an additional interpretation. In view of the fact that the difference between the median scores of fourth grade whites and fourth grade negroes in rhythm is 2.79, a comparatively large difference, a question as to the possible cause of this difference undoubtedly arises. A tentative interpretation would suggest one of two causes for this large difference. On the one hand, if the median chronological age of fourth grade negroes was higher than that of fourth grade whites, the factor of maturity might serve as a possible explanation. On the other hand, if the statistical results concerning the median chronological age of all other grades were found to be similar, relatively, to the results obtained in regard to the fourth grade, the factor of statistical unreliability might serve as a possible explanation. Since the median chronological age of fourth grade negroes has been found to be ten years and the median chronological age of fourth grade whites, nine years, the one year advance in chronological age of the negroes over the whites may serve to suggest that the explanation of the results obtained is based upon the factor of the greater maturity of the fourth grade negro as compared with that of the fourth grade white. However, the writer is fully aware of the fact that this interpretation is merely tentative since a more thorough investigation of the relationship between chronological age and rhythm might suggest statistical unreliability as a cause of the comparatively large difference between the medians for fourth grade children

of both races. (See page 70, suggestions for further investigation, no. 4.)

IX. SUGGESTIONS FOR FURTHER INVESTIGATION

Studies concerning any one of the phases of musical ability—pitch, intensity, time, consonance, rhythm, and tonal memory—may be treated by various methods of investigation, some of which are suggested in the following:

1. Studies of primary interest:
   (1) A comparison of white and negro children as a whole.
   (2) A comparison grade for grade of white and negro children.
   (3) The correlation between intelligence and any one phase of musical ability.
   (4) The correlation between maturity and any one phase of musical ability.
   (5) The relationship between any one phase of musical ability and intelligence and this same phase and maturity.
2. Studies of secondary interest:
   (1) A comparison of white and negro boys.
   (2) A comparison grade for grade of white and negro boys.
   (3) A comparison of white and negro girls.
   (4) A comparison grade for grade of white and negro girls.
   (5) A comparison of white boys and white girls.
   (6) A comparison of negro boys and negro girls.

REFERENCES

(1) Ferguson, G. O.: The psychology of the negro. Archives of Psychology, 1916, Vol. 5, No. 36.
(2) Garth, T. R.: White, Indian, and negro work curves. Jour. Applied Psych., 1920, Vol. 5, pp. 14–25.
(3) Hirsch, N. D.: A study of natio-racial mental differences. Genetic Psych. Monog., 1926, Vol. 1, Nos. 3 and 4.
(4) Murdock, K.: A Study of race differences in New York City. School and Society, 1920, Vol. 11, pp. 147–150.
(5) Peterson, J., and Lanier, L. H.: Studies in the comparative abilities of whites and negroes. Mental Measurement Monographs, February, 1929, No. 5, Part II.
(6) Peterson, J.: Comparative abilities of white and negro children. Comp. Psych. Monog., 1923, Vol. 1, No. 5.
(7) Peterson, J.: A functional view of consonance. Psychol. Rev., 1925, Vol. 32, pp. 17–33.

(8) SEASHORE, C. E.: Manual of instructions and interpretations for tests of musical talent. Columbia Phonograph Company.
(9) SEASHORE, C. E.: The psychology of musical talent. Boston, New York, Silver, Burdett and Co., 1919.
(10) SUNNE, D.: Comparative studies of white and negro children. Jour. Applied Psych., 1917, Vol. 1, pp. 71–84.

STUDIES OF TYPOGRAPHICAL FACTORS INFLUENCING SPEED OF READING

V. Simultaneous Variation of Type Size and Line Length[1]

MILES A. TINKER and DONALD G. PATERSON

University of Minnesota

In our search for typographical arrangements favoring maximum speed of reading we have found: (1) 10 point type in comparison with 6, 8, 12, and 14 point type is the optimum size of type *for a line length of 80 mm. or 19 picas*, and (2) 80 mm. or 19 pica line length (or width) in comparison with 59, 97, 114, 136, 152, 168, and 186 mm. line lengths is the optimum *for 10 point type size*.[2] The question naturally arises as to what would be the effect of a simultaneous variation of type size and line length. In such an experiment both size of type and length of line would be varied simultaneously, the variations from the standard being printed line for line, i.e., with the words in each line of each copy of the text exactly the same. Such a study would show whether optimal conditions, as far as these two factors are concerned, are a resultant of both factors working together or whether these factors are independently variable in their effects. If the former be true, then we might expect to find

[1] The writers are grateful to the following for aid in conducting this study: J. B. Lippincott Co., Philadelphia, for permission to reprint the Chapman-Cook Speed of Reading Tests, Forms A and B, for experimental purposes; the University of Minnesota Graduate School for research grant; and the instructors in charge of the Freshman English classes in the College of Engineering, University of Minnesota, for permission to test their students.

[2] Donald G. Paterson and Miles A. Tinker, Studies of Typographical Factors Influencing Speed of Reading, II. Size of Type, J. Appl. Psychol., Vol. XIII, No. 2, April, 1929, pp. 120–130; III. Length of Line, Vol. XIII, No. 3, June, 1929, pp. 205–219.

several combinations of both variables which would be equally efficient in promoting rapid reading.

The particular problem before us has been ignored by previous experimenters with the exception of Gilliland.[3] He varied size of type and length of line by photographically enlarging and reducing a standard short paragraph which had been set in 12 point type with a line length of 3½ inches (85 mm. or 20 picas). In our review of the literature on the Size of Type article we mentioned a number of points definitely limiting the significance of his findings. At this time, we would add that the method of photographic enlargement and reduction does not duplicate printing conditions, since printing a given piece of text in different type sizes, *line for line*, i.e., keeping the words in each line of each copy of the text exactly the same, does not result automatically in a proportional change in heighth and width of letters and spaces such as occurs in the photographic process. Hence, the practical significance of Gilliland's findings would necessarily be limited, since they could not be applied directly to printing practice.

PROCEDURE[4]

In this experiment we took the previously ascertained optimal line length of 80 mm. or 19 picas for 10 point type as a standard (Chapman-Cook Speed of Reading Test, Form A, set up in lower case, 10 point Scotch Roman, 19 pica line length, set solid) and compared efficiency in reading this standard with efficiency in reading Form B of the Chapman-Cook Test set *line for line* in 6 point, 8 point, 12 point, and 14 point type. This procedure yielded line lengths for Form B as follows: for 6 point type, 16 picas; 8 point, 17 picas; 12 point, 23 picas; and 14 point, 27 picas. The following five test groups of 80 students each were used (group III was introduced as a control group to

[3] A. R. Gilliland, The Effect on Reading of Changes in the Size of Type. El. Sch. J., 1923, 24, 138–146.

[4] For details regarding general printing of test materials and methodology see previous articles in this series, especially, II. Size of Type.

demonstrate again that Forms A and B are approximately equal in difficulty when typography of the two forms is identical):

Group I. Form A, 10 pt., 19 picas, Form B, 6 pt., 16 picas
Group II. Form A, 10 pt., 19 picas, Form B, 8 pt., 17 picas
Group III. Form A, 10 pt., 19 picas, Form B, 10 pt., 19 picas
Group IV. Form A, 10 pt., 19 picas, Form B, 12 pt., 23 picas
Group V. Form A, 10 pt., 19 picas, Form B, 14 pt., 27 picas

In all, 400 students in Engineering Freshmen English classes at the University of Minnesota were tested. None of these students had previously been exposed to the Chapman-Cook Tests. Engineering freshmen were used as subjects because available psychology students had already taken these tests.[5] The classes tested were allotted at random to each of the writers who conducted all of the tests in accordance with our previous practice as regards testing procedure and time-limits.

RESULTS

Confining attention to one of the test group comparisons at a time will facilitate the interpretation of the results shown in table 1. For test group I, column three shows that the average number of paragraphs read was 16.26 for Form A, 10 point, 19 pica line length, and was 15.14 for Form B, 6 point, 16 pica line length. This is a difference of 1.12 paragraphs (column 6) and indicates that the standard Form A is read 7.4 per cent faster than the 6 point type in Form B (column 8). That this is a statistically certain difference is shown in column 10. Data on variability, essential for deriving the values in columns 9 and 10, are shown in columns 4 and 5. It is apparent that reduction in type size from 10 point to 6 point, even though printed line for line which results in proportional reduction in line length, tends definitely to slow down speed of reading.

The results for test group II indicate that there is no marked

[5] Students once exposed to these tests are no longer suitable subjects for the type of investigation here reported. See Donald G. Paterson and Miles A. Tinker, IV. Effect of Practice on Equivalence of Test Forms. J. Appl. Psychol., Vol. 14, No. 3, June, 1930, 211–217.

TABLE 1

Influence of simultaneous variation in size of type and length of line

Differences given are for the mean score on Form A, 10 point, 80 mm. or 19 pica line minus the mean score on Form B, varying in type size and length of line (6 point, 16 picas; 8 point, 17 picas; 10 point, 19 picas; 12 point, 23 picas; 14 point, 27 picas). All test forms printed in Scotch Roman set solid on egg shell paper stock. The mean score is the average number of paragraphs of 30 words each in the Chapman-Cook Speed of Reading Tests read in $1\frac{3}{4}$ minutes. In each sub-group, N = 80 freshmen in College of Engineering English classes, University of Minnesota.

TEST GROUP	TEST FORM, TYPE SIZE AND LENGTH OF LINE	MEAN	P.E.$_{dist.}$	P.E.$_{M}$	DIFFERENCE BETWEEN MEANS IN			P.E.$_{diff.}$*	$\frac{D}{P.E._{diff.}}$
					Paragraphs	Words per second	Per cent		
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
I	A, 10 pt., 80 mm. or 19 picas	16.26	2.79	.31	1.12	0.32	7.4	.19	6.01
	B, 6 pt., 68 mm. or 16 picas	15.14	2.47	.27					
II	A, 10 pt., 80 mm. or 19 picas	15.84	3.06	.34	0.25	0.07	1.6	.18	1.35
	B, 8 pt., 72 mm. or 17 picas	15.59	2.63	.29					
III	A, 10 pt., 80 mm. or 19 picas	16.18	3.01	.34	0.15	0.04	0.9	.17	0.88
	B, 10 pt., 80 mm. or 19 picas	16.03	2.47	.28					
IV	A, 10 pt., 80 mm. or 19 picas	17.10	2.91	.32	0.67	0.19	4.1	.19	3.56
	B, 12 pt., 97 mm. or 23 picas	16.43	2.38	.26					
V	A, 10 pt., 80 mm. or 19 picas	16.64	2.62	.29	0.70	0.20	4.4	.19	3.64
	B, 14 pt., 115 mm. or 27 picas	15.94	2.26	.25					

*$PE_{diff.}$ values in column 9 were computed on the basis of the standard formula for determining the statistical significance of obtained differences for *correlated measures*. The correlations for the five sub-groups in this experiment were $+0.80 \pm .03$ for group I; $+0.84 \pm .02$ for group II; $+0.86 \pm .02$ for group III; $+0.81 \pm .03$ for group IV; and $+0.76 \pm .03$ for group V.

difference in speed of reading 10 point set in 19 pica line lengths and 8 point set in 17 pica line lengths. The difference of 0.25 paragraphs (column 6) is slight and statistically uncertain (column 10). For all practical purposes we would conclude that it is possible to reduce type size from 10 point to 8 point without loss of reading efficiency providing this reduction is accompanied by a proportional reduction in line length. This negligible difference is in marked contrast to the 5.2 per cent loss in reading speed which we found when type size was reduced from 10 point to 8 point keeping line length constant at 19 picas. Thus it is probable that within definite limits type size may be increased or decreased if accompanied by proportional increases or decreases in line length without influencing speed of reading.

As was to be expected, test group III shows no significant difference in reading performance since Forms A and B were typographically identical. The rôle of such a control group is important because results secured from the experimental procedures employed in this series of investigations are subject to sampling errors. Furthermore, in view of the short time-limits used, it is quite likely that differences in motivation and even in small details of test administration may upset equivalence in difficulty in the two test forms. By inserting a control group in each experiment it is possible to detect any subtle irregularities that might tend to invalidate results.

Test groups IV and V yield almost identical results. There is about a four per cent differential in favor of the standard 10 point set-up. The differences, however, are not statistically certain, although probability points in that direction. This lack of any very marked decrease in reading performance is not so surprising for group IV where 12 point type was used but did seem somewhat striking for group V where 14 point type was used. In the latter case there is a marked increase in length of line (from 19 picas for the standard 10 point to 27 picas). This increase in length of line and the increase in type size too necessitated the use of a very much larger printing space. The outside measurements of the printing area in the standard, Form A, were $6\frac{10}{16}'' \times 8\frac{9}{16}''$ whereas the same measure-

ments for Form B (14 pt., 27 pica line length) were $9\frac{4}{16}'' \times 12\frac{3}{16}''$. This is an area difference of approximately one to two. Apparently, size of type may be increased markedly providing a proportional increase in line length is introduced simultaneously. This finding is significant not only from the viewpoint of the printer and advertiser, but also from the standpoint of theoretical psychology in its consideration of perception in reading.

It is unfortunate that our typographical arrangements were not extensive enough to indicate the point where increase in size of type significantly slows down reading performance. The printing facilities at our disposal precluded the preparation of tests printed in type larger than 14 point. Such extension of the work should be undertaken in the interests of advertising theory and practice.

From a theoretical point of view, the results obtained in this experiment tend to confirm our previous views regarding the probability that reduction in legibility of the printed characters accompanying reduction in the size of the letters (from 10 point to 6 point type) is a factor of importance in retarding speed of reading. However, the present results do not confirm our previous views regarding the supposed reduction in legibility of 8 point type as compared with 10 point, both set in 19 pica line lengths. When 8 point type is set with 17 pica line length as in this experiment, it is read at practically the same rate as 10 point type set with 19 pica line length.

This discrepancy in the results with 8 point type raises an interesting and important question concerning perception in reading. This question has to do with the rôle of contextual sequences in providing premonitions of meaning which favor rapid reading. It would seem as though the maintenance of the spatial relationships between the words in successive lines of print which is accomplished by the *line for line* arrangement irrespective of increase in size of type permits equally efficient reading performance to take place with 8 and 10 point type[6] and

[6] Since 9 point type is intermediate in size, it would seem reasonable to assume that the findings for 8 and 10 point type would also hold for 9 point.

possibly with larger sizes. This emphasizes the importance of sentence arrangements which will allow a maximum utilization of vertical cues favoring premonitions of meanings.

SUMMARY AND CONCLUSIONS

1. In continuing previous work of the writers on size of type and length of line, the present experiment varied both factors simultaneously. This was accomplished by printing the variations *line for line* using 10 point type with a 19 pica line length as a standard. The Chapman-Cook Speed of Reading Tests were printed in 6, 8, 10, 12, and 14 point type with line lengths of 16, 17, 19, 23, and 27 picas respectively. There were 400 college freshmen used in the experiment, none of whom had previously taken these tests.

2. The text printed in 6 point type with 16 pica line length was read more slowly than the standard text printed in 10 point type with 19 pica line length. The difference of 1.12 paragraphs means that the standard text was read 7.4 per cent more rapidly.

3. The text printed in 8 point type with 17 pica line length was read at approximately the same rate as the standard text.

4. Texts printed in 12 point type with 23 pica line length and in 14 point type with 27 pica line length were read at a slightly slower rate than the standard text.

5. The results obtained in this experiment seem to justify the conclusion that *line for line* arrangements varying from a standard text printed in 10 point with an optimal line length of 19 picas yield equally efficient reading performance for 8 and 10 point type, and possibly for larger sizes. The evidence shows that 6 point type with line length reduced as stated is read at a significantly slower rate.

6. The experiment yields evidence that, within certain limits, neither size of type nor line length as single variables can be relied upon in determining optimal typographical arrangements. Both factors (and perhaps others as well) work hand in hand and must be properly balanced to produce a printed page which will promote a maximum reading rate.

THE RELIABILITY OF THE LINCOLN HOLLOW SQUARE FORM BOARD, AND A COMPARISON OF HOLLOW SQUARE SCORES WITH STANFORD BINET MENTAL AGES

EDWARD A. LINCOLN

Harvard University

When the original description of the Lincoln Hollow Square Form Board was published[1] there were not at hand sufficient data for studying the reliability of the board. Out of about 500 reports of cases sent in to the writer, 139 were in such form as to be useful for a reliability study. For many individuals, the Stanford Binet mental ages were reported along with the Hollow Square scores, and thus data were provided for an examination of the relationship between the performances on the two tests.

The Hollow Square test has eight parts or problems, each of which is timed separately. It was therefore easy to study the reliability of this test by the split item method, using the total time on the odd numbered problems for one variable, and the total time on the even numbered problems for the other, and correcting the obtained coefficient by the Spearman-Brown formula. Since the number of cases was small, it was not possible to study the reliability of the test at different chronological ages. The reliability coefficient for the whole group was $.952 \pm .007$, which is higher, of course, than that which would be obtained in any single age group. The mean chronological age of the 139 cases was 108.7 months, with a standard deviation of 33.8 months. The mean total time for the group was 235.6

[1] Dearborn, W. F., A Series of Form Board and Performance Tests of Intelligence Tests. Harvard Monographs in Education, No. 4. Graduate School of Education, Harvard University, Cambridge, Mass.

seconds, with a standard deviation of 93.9 seconds. These figures will be useful if further reliability studies are made on groups which have different variabilities.

TABLE 1

Correlations in different chronological age groups between the Stanford Binet mental ages and hollow square scores

CHRONOLOGICAL AGES	r, M.A. TIME	MENTAL AGE IN MONTHS		TIME IN SECONDS		N
		Mean	S.D.	Mean	S.D.	
9	.53 ± .07	94.3	18.6	208.4	79.6	44
8	.60 ± .05	85.6	17.4	254.7	78.0	69
7	.58 ± .06	80.2	16.8	262.7	70.7	60
6	.25 ± .09	70.3	12.2	306.9	65.9	52
5	.49 ± .08	60.8	10.6	326.8	51.1	31
4	.60 ± .09	48.5	8.5	352.3	46.5	26
All	.64 ± .02	76.9	20.4	275.8	81.9	282

TABLE 2

Differences between the Stanford Binet mental ages and performance ages on the Lincoln hollow square

Difference in months

CHRONOLOGICAL AGE	0-4	5-9	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49	50+	N	MEDIAN	Q_1	Q_3
9	7	4	5	3	5	3	3	3	1	2		36	18.3	7.5	30.0
8	12	11	9	6	6	5	9	3	1	3	1	66	15.8	7.0	30.3
7	11	11	12	7	6	5	4	1				57	12.7	6.5	23.5
6	12	7	4	10	3	4	5	4	1	1	1	52	16.5	5.7	28.8
5	6	9	7	4	1	2	1	1				31	10.4	6.0	18.4
4	6	5	8	1	3	2	1					26	11.3	5.5	17.5
3	6	4	4	1	1	1						17	8.1	3.6	13.4
2	3	1	0	1	0	1						6	5.0	2.5	17.5
All	63	52	49	33	25	23	23	12	3	6	2	291	13.1	5.9	24.3

The comparison of Hollow Square scores with Stanford Binet mental ages was first made by use of correlation coefficients. Table 1 shows the obtained coefficients for chronological ages 4

to 9, inclusive, together with other important data covering the groups. It will be noted that the relationship is not a close one, even in the group composed of all ages combined. The indication is that a coefficient of about .50 represents the relationship in any single chronological age group.

This relationship was also studied in another way. By reference to the standards,[2] the time scores on the Hollow Square were transmuted to "performance ages," and the latter were compared with the obtained mental ages. Table II shows the distributions of differences, taken without regard to direction, between these two ages. Here again the evidence is substantial that the relationship is not close. Except in ages two and three, where the numbers of cases are too small to make the results significant, the median difference is practically a year, at least. In only about a quarter of the cases are the differences small enough to be disregarded. It is very clear from these figures that the Hollow Square is measuring something substantially different from the ability or abilities measured by the Stanford Binet scale. This is, of course, a common finding in regard to the relationship between scores on the Stanford Binet scale and performance or form board tests.

SUMMARY

1. A reliability coefficient of .952±.007 was found for the Lincoln Hollow Square Form Board using a group with rather large variability.

2. Like other performance tests, the Hollow Square does not give scores closely related to Stanford Binet mental ages, the coefficients averaging around .50 in separate chronological age groups.

[2] Revised standards were published in the Journal of Applied Psychology, Vol. XI, No. 4, August, 1927, p. 264.

MOTION STUDY TABLES FOR CLERICAL OPERATORS

HERBERT W. ROGERS

Lafayette College

Poffenberger's recent book[1] brought to this author's attention the fact that no detailed reports of the results obtained by motion study have appeared in the psychological literature since the early work of Gilbreth.

Twelve years ago four motion study tables were made for clerical operators at a mail order concern, the Charles William Stores in Brooklyn, N. Y. Results of the operations on two of these tables are at hand. Unfortunately photographs of these two tables are not available; hence they have been sketched from memory.

Since the concern was not ready to meet the necessary expenditure no motion pictures for the accurate study of the motions of the operators were taken at old or new tables. The motions were, therefore, but crudely observed. The tables were designed so that the materials for work would be placed conveniently. They were made so that the worker could either sit or stand and at once be at the same relative position in respect to the table. The tables were made for specific workers. It would be possible, however, by means of a sliding groove in the legs of the tables and chairs, to make them easily adjustable for any girl.

Table 4 was made for the work of enclosing letters in envelopes.

The work to be performed on table 4 was simply enclosing letters. Material to be manipulated, a letter with appended envelope, was placed in the receptacle labelled "A." This had

[1] Applied Psychology, Appleton, 1928.

a free sliding bottom to which was attached a spring so that the letter to be taken was always just a little above the level of the table. Receptacle "B," for the letters enclosed in envelopes, also had a spring bottom so that the envelopes would rest in an orderly manner and not topple over. The rear of this box

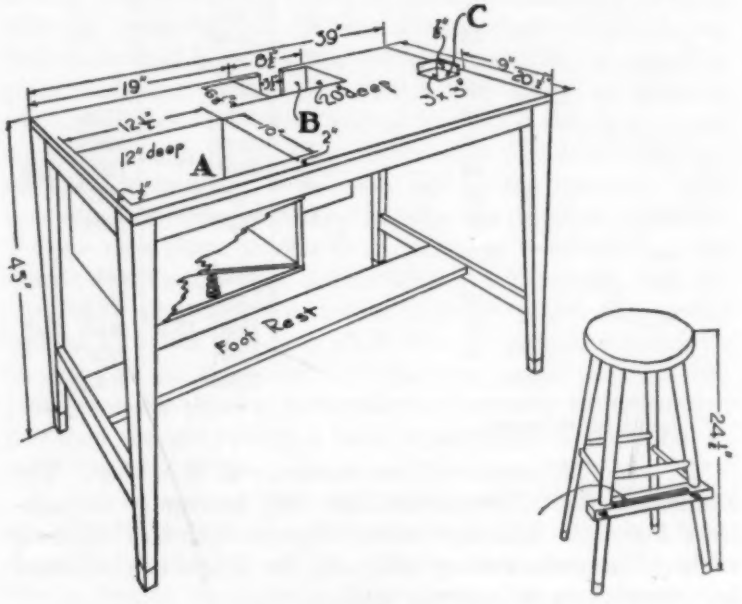

FIG. 1. TABLE 4 FOR ENVELOPING LETTERS

A, to hold letters; *B*, to hold enveloped letters; *C*, to hold paper clips taken from material. (*B* has spring as shown in section of *A*.) One-half-inch elevation to grip heels.

was removed so that a hand could fit in, on top and bottom, to lift out the column of envelopes. Box "C" was for clips taken from the letters.

The following table shows the results of the operation on this table:

Motion Study Table No. 4

	WEEK ENDING	ENCLOSURES PER 8½ HOUR DAY			
		L		M	
(8 weeks)	Nov. 4	480		502	
(3 weeks)	" 25	912		1043	
	Dec. 2	900	955	*	
	" 9	998		*	
	" 16	1009		*	
	" 23	*1347*		*	
	" 30	*1322*		1134	
	Jan. 6	*1363*	*1416*	1280	
	" 13	*1492*		1580	1480
	" 20	*1425*		1717	
	" 27	*1545*		1690	
	Feb. 3	1343		*1956*	
	" 10	1357		*1884*	
	" 17	1391	1319	*1643*	*1848*
	" 24	1292		*1632*	
	Mar. 3	1212		*2124*	

* Absent for sickness.

"L" and "M" represent two operators at this work. The numbers in the table represent the daily average of two subjects for each indicated week. Figures italicized indicate work on the motion study table and the figures not italicized indicate work on the common table.

When the investigtor went to work previous records of output were at hand. These appear in the first row of the table. Things were going very easily in this department and no incentive for a fair day's work was provided. Starting with the week of November 6, a bonus system was instituted. Each girl whose output was above the average received, in addition to her daily wage, a percentage of her wage equal to the percentage above average of her output for that day. For instance, if the output was as follows:

A.	150	+50%
B.	100	00
C.	50	−50%

A's output was 50 per cent above the average; therefore if she earned $7 a week or $1.16 a day, she received $.58 extra this day. The table shows the effects of such a system of reward for superior accomplishment. In three weeks output was doubled. This sort of a bonus is thought to be superior to most other means of paying a bonus. It is immediate in that the results of the previous day's work are posted the following morning with the bonus figured. It lacks the drawback of most bonus systems in that in most instances one guesses at a standard of accomplishment, (sometimes after a time study, on subjects who due to lack of a reward are not striving as they might) and pays the bonus to all who exceed that standard. By our method the standard is set each day by the workers. This is an excellent means of finding out what the standard should be. When such a bonus scheme is in operation, however, the time shortly approaches when the variability of the group, which is constantly decreasing, comes to a point where the actual amount of bonus paid is so slight that its operation ceases to be much of an incentive. At this point it is evident that production, for existing conditions, is about at a maximum and it is then necessary to set a fixed standard and to pay a bonus to all who exceed this.

It will be noticed that each of the girls, "L" and "M," alternated periods on the motion study table with the periods on the old type of table and that these periods seem to be sufficiently long to demonstrate their value. The results seem to show very definitely that more work is accomplished on the motion study table than on the old type table. Each girl was asked which table she preferred to use and each said the motion study table.

At the common type of table "L" had an average output of 1319 enclosures per day. She received $7 a week or $1.16 a day. The cost per enclosure was $.00088. At the motion study table she averaged 1416 enclosures per day; $.00082 per enclosure; an improvement of 97 a day or a saving of $.085 a day or $.51 a week. At the common type of table "M" had an average of 1480 enclosures a day. As she received

$8 a week or $1.33 a day, the cost per enclosure was $.00090. At the motion study table she averaged 1848 enclosures per day ($.00072 per enclosure) an improvement of 368 a day or a saving of $.3307 a day or $1.98 a week.

The cost of the sample table was $13.50. The estimated cost of 3 more tables was $11 each. These prices were much too high. The reason for this being that the boss carpenter thought plain tables would be just as good; could see no good reason why directions for construction had to be so specific; thought that they were too high from the floor, etc. He was most unwilling to make them. "L" will then pay for a table in 22 weeks and "M" will pay for her's in 5½ weeks. The two tables would pay for themselves in 14 weeks and beginning with the 15th week there would be a saving of $2.49 per week or 17 per cent on these two girls.

If one wishes to be precise then, to this saving one must add the floor space saved; the expense of employing and maintaining the girls who would be eliminated by the increased output of the others. There were 16 girls in this division, with motion study tables for all this number would be reduced to 13; the cost of welfare work, etc., money lost by absences, etc., would probably raise the saving at least 10 per cent which makes the ultimate saving about 27 per cent.

Motion Study Table 1 was made for "mail openers," this work consisted of opening letters and inspecting remittances.

Incoming mail was put through a machine which slit open the top of the envelope. "A" was a slide on which were placed the envelopes which had the tops slit. They slid down so that there was always an envelope at the end of the slide in a position to be grasped easily. The contents of the envelope were removed to the flat space of the table and the envelope thrown in box "F." The letter was stamped, with the rubber stamp at "C," the money counted and a notation made, with pencil, of the kind of money enclosed; such as silver, bills, check, money order, or stamps, in a space provided by the stamp. If the money was in specie, it was put in an envelope taken from "E," which was inclined at an angle of 45 degrees.

When finished with the order, the operator threw it into "B" which had a floor set with a spring so that the letters would remain in an orderly pile.

The work in this department was figured on an hourly basis for it was seldom that there were more than 4 to 6 hours work

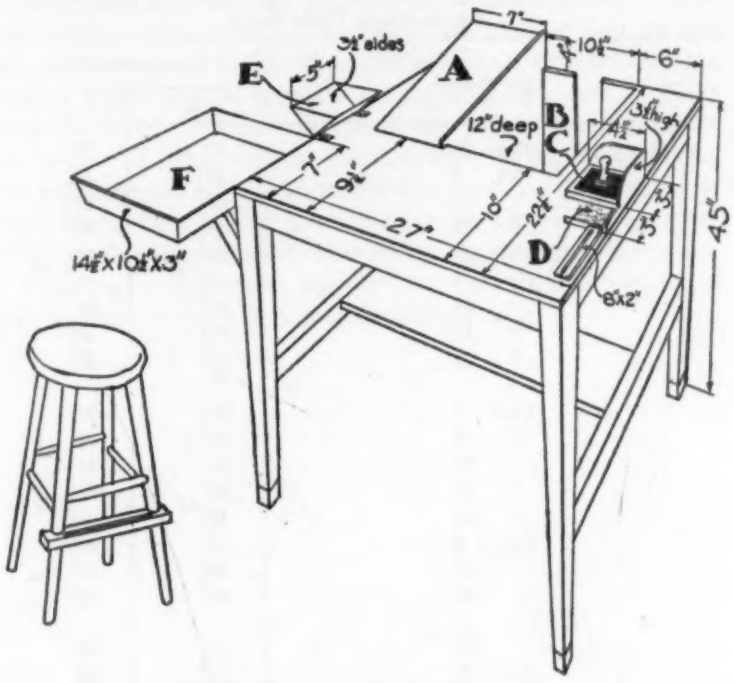

FIG. 2. TABLE 1 FOR MAIL OPENERS

A, envelopes, cut open at top; *B*, orders taken from envelopes; *C*, rubber stamp; *D*, pins; *E*, money envelopes; *F*, basket for waste envelopes.

per day for an employee. These girls appeared an hour and a half earlier in the morning than any other group in order that the mail could be opened and the work ready for the next department. The short day, on this kind of work, was because all the mail had to be opened as near 11 a.m. as possible for proper routing to other departments.

The same variety of bonus was paid here as in the enclosing department. This bonus started on the week beginning October 2.

The following table shows the results which were obtained from the operations at this table:

Motion Study Table No. 1

WEEK ENDING		LETTERS PER HOUR					
		G			N		
(2 weeks)	Sept. 30	39			27		
	Oct. 7	*45*			30		
	" 14	*49*			32		
	" 21	*46*			30		
	" 28	*49*	*49*		29	30	
	Nov. 4	*49*			31		
	" 11	*51*			28		
	" 18	*50*			30		
	" 25	*52*			31		
	Dec. 2	43			*35*		
	" 9	40			*39*		
	" 16	43			*38*		
	" 23	44	42		*39*	*39*	
	" 30	43			*41*		
	Jan. 6	43			*40*		
	" 13	42			*39*		
	" 20	41			*41*		
	" 27	*48*			30		
	Feb. 3	*51*			31		
	" 10	*51*			29		
	" 17	*53*	*52*		31	31	
	" 24	*54*			32		
	Mar. 3	*52*			30		
	" 10	*51*			33		
	" 17	*53*			33		

At the common type of table "G" had an average output of 42 per hour and at the motion study table an output of 52 per hour. This was a saving of 24 per cent. At the common

type of table "N" had an average output of 31 per hour and at the motion study table an output of 39 per hour. This was a saving of 26 per cent. To these savings may be added a 10 per cent saving as with the motion study table for envelopers which gives us an average saving of 35 per cent.

Estimated cost of more tables such as this was $16. Each girl received $8 a week. If we figure the working day, on this kind of work, at 5 hours, then the operators received $4.71 a week. Since they averaged a saving of 25 per cent or $1.18 each per week, two new tables would pay for themselves in 14 weeks. Adding to this the overhead saved, 10 per cent, the tables would pay for themselves in 10 weeks and thereafter there would be a saving of $1.65 on $4.71 a week. The $4.71 is the weekly wage for a five hour day on this kind of work.

From the foregoing it appears that this crude variety of motion study results in a considerable saving.

A NOTE CONCERNING RHYTHM TESTS IN POETRY AND IN MUSIC

MARION HAMMOND

San Diego State College

The purpose of this study was to ascertain the degree of relationship that exists between an individual's sense of rhythm in poetry and his sense of rhythm in music as measured by the Logasa-McCoy-Wright Tests for the Appreciation of Literature (Test 6) and Part 6 (Rhythm) of the Seashore Tests of Musical Accomplishment.

The data were assembled in two classes of Educational Measurement in San Diego State College during the fall semester of 1929–1930. The two series of scores are summarized in table 1.

TABLE 1

	N	RANGE	MEAN	σ	C.V.	r
Logasa-Wright.....	84	0–8	4.35 ± .12	1.63	37.64	
Seashore.....	84	34–96	73.19 ± .79	10.70	14.62	.18 ± .11 (84)

The mean for the distribution of scores of the poetry rhythm test is 4.35 ±.12 as compared with 4.9, the norm for college juniors. The mean of the music rhythm test is 73.19 ±.79 as compared with 74.5, the norm for adults.

The coefficient of correlation is .18. While the correlation is positive it is not large enough to be at all significant. The low correlation leads to two possible conclusions; either the sense of rhythm in poetry and the sense of rhythm in music are entirely independent of each other; or, one or both of the tests used are not valid, that is they do not measure that which they purport to measure.

To the writer the nature of the two tests seems to indicate that the validity of the Seashore test is much higher than the

validity of the Logasa-McCoy-Wright test, and that the probable lack of validity of the Logasa-McCoy-Wright test is one of the main factors affecting the correlation.

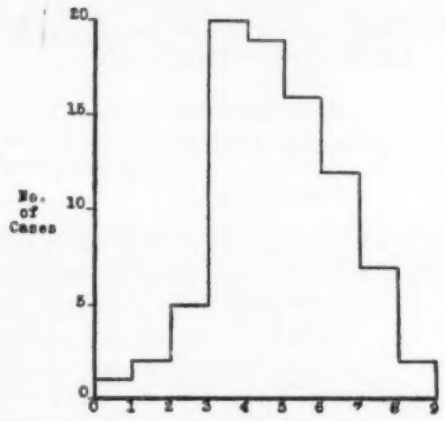

FIG. 1. DISTRIBUTION OF SCORES FOR LOGASA-MCCOY-WRIGHT RHYTHM TEST

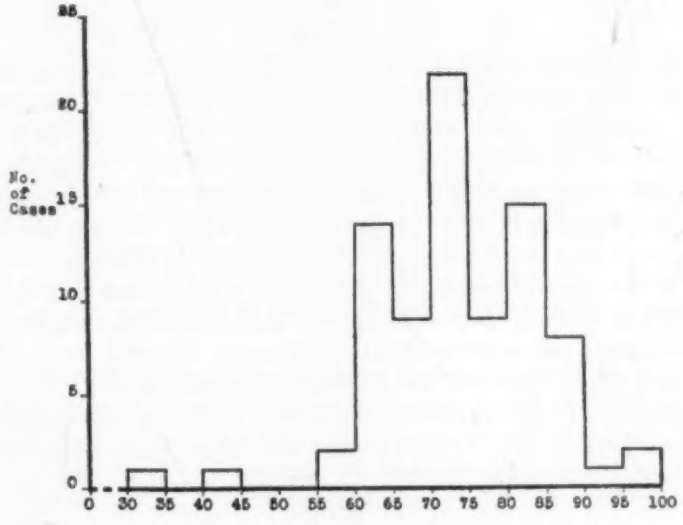

FIG. 2. DISTRIBUTION OF SCORES FOR SEASHORE RHYTHM TEST

THE LANGUAGE DIFFICULTY OF SPANISH-AMERICAN CHILDREN

B. F. HAUGHT

University of New Mexico

INTRODUCTION

Teachers have found that Spanish-Americans do school work that is very much inferior to that done by Anglo-Americans. When the teachers are asked why this condition exists, they usually attribute it to an inability to use and understand the English language. When intelligence tests are administered to both groups, the children of Spanish descent fall considerably below the standards obtained by those of Anglo descent. There is an inclination to assume that this does not mean an inferiority, but a language difficulty encountered in taking the tests.

THE PROBLEM

The purpose of this study is to find some objective evidence that such a language difficulty does or does not exist. It seems reasonable to assume that a language handicap encountered in taking an intelligence test should decrease as the children become older, progress through the grades, and thus become better acquainted with the English language. This should be the case especially when the children are in schools where English is used altogether in the classes. If this handicap does decrease with age, the intelligence quotients should increase accordingly and approach that of the Anglo children.

This study compares the intelligence quotients with chronological ages. All the Spanish children from four school systems were selected. They range in ages from seven to nineteen, and in grades from one to twelve.

THE METHOD

The Pintner-Cunningham Mental Test was used in grades 1 to 3; the National Intelligence Tests in grades 4 to 7; and the Terman Group Tests of Mental Ability in grades 8 to 12. Pupils in grades 7 to 12 whose mentality was too low to be

TABLE 1

Showing the distribution of subjects on the basis of intelligence quotients and chronological ages. Circles through which the line passes represent the means of the intelligence quotients for the successive ages

Intelligence Quotients

Chronological Ages in Months	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80-84	85-89	90-94	95-99	100-104	105-109	110-114	115-119	120-124	
229-220			2	5	3	4	8	7	9	2	3	1						44
219-210						1	6	4	3	2	1	1						18
209-200						5	9	7	7	4	5	1						38
199-190		1	2		2	11	7	8	6	4	3	1		2				47
189-180		1	3	1	5	3	9	7	5	11	5	2	2					54
179-170	1	4	3	6	8	5	3	8	7	6	4	3	1	1				60
169-160			3	4	6	12	5	8	9	11	6	2	1	1	4	1		73
159-150		3	1	4	13	12	10	14	13	6	9	2	4	2				93
149-140			5	5	3	7	13	13	6	8	7	2	1	2		1	1	74
139-130			2	4	3	5	15	7	9	3	2	7	3	3		1		64
129-120	1		2		8	9	7	12	8	2	5	1	2	2	1		2	62
119-110			1		3	2	6	6	7	8	7	5	3	1	1		1	51
109-100			1		5	3	4	5	4	6	12	3	3	2				48
99-90		1	2	3	2	1	2	8	8	4	2	3	2	4			1	43
89-80			3	1	3	5	1	1	2	6	3	4	1	2	2			34
79-70				2	3	1	1	2	2	1	1	4	2		1		2	22
	2	10	30	35	67	86	106	117	105	84	75	42	25	22	9	3	7	825

measured by the Terman Group Tests of Mental Ability were given the National Intelligence Tests. Pupils in grades 4 to 7 whose mentality was too high to be measured by the National Intelligence Tests were given the Terman Group Tests of Mental Ability, and those whose mentality was too low were

given the Pintner-Cunningham Primary Mental Test. Some children in grades 1 to 3 were found to have a higher mental age than could be measured by the Pintner-Cunningham Primary Mental Test. These were given the National Intelligence Tests.

The tests were scored and the raw scores converted into intelligence quotients. The correlation between intelligence quotients and chronological ages was found statistically and is also shown graphically.

Table I presents no evidence that intelligence quotients increase with age and school attendance. On the whole there is a slight decrease, due mostly to a sudden drop at about ten years of age. This sudden drop is difficult to understand. It may be due to some factor of selection that came from combining the pupils from several schools. It is too high to mark the dividing line between the use of the Pintner-Cunningham and the National Intelligence Tests and too low to mark the transition from the National Intelligence Tests to the Terman Group Tests.

The general appearance of the scatter-diagram indicates a negative correlation between chronological ages and intelligence quotients. The statistical correlation (Product-Moment method) is $-.013$. This is practically zero, which one would expect if intelligence quotients remain constant.

CONCLUSIONS

1. Intelligence quotients of Spanish children as revealed in this study do not increase with chronological age, but there is a slight tendency for them to decrease. This conclusion assumes that no selective factor is involved, and it is likely that selection would operate to increase rather than decrease the intelligence quotients.

2. There is a sudden drop in intelligence quotients at about ten years of age.

3. Since the older children are handicapped as much as the younger, there seems to be no justification for assigning the difficuty to inability to use or understand English.

4. Until the existence of a language handicap on the part of Spanish children is demonstrated by properly controlled experiments, it seems safer to avoid using the concept as an explanatory principle in educational problems.

5. The average Spanish child has an intelligence quotient of 79 as compared with 100 for the average Anglo child.

6. There are some Spanish children as bright as the very superior Anglo children.

PERSONALITY TRAINING AND THE CHANGING EMPHASIS IN THE LIBERAL ARTS COLLEGE

G. H. ESTABROOKS

Colgate University

At the present day, the liberal arts college is subject to a great deal of criticism from various sources. Needless to say, much, perhaps indeed the majority of this criticism is uncalled-for and quite beside the point. Someone has said that there are more tinkers than thinkers in higher education. While we admit the fundamental justice of this statement, we can not close our eyes to the wide-spread dissension which seems to exist concerning the four-year liberal arts college.

The protests against this institution are assuming various and interesting forms. Perhaps the most remarkable of these is the Junior College, which is literally threatening the life of the four-year institution in some of the Western states. There can be no doubt but that the movement will become more general and widespread in the near future.

The semi-professionalizing of the liberal arts course is another protest; while the interesting and significant experiments at Harvard and Swarthmore, seeking as they do to approximate the Oxford ideal, represent perhaps the most outstanding modification taking place within this type of college itself.

The purpose of this paper, however, is to concentrate attention upon another phase of remodification, the need for which has been brought clearly home to the writer through his association with personnel and vocational guidance problems in his own institution.

It is now merely a truism for one to assert that the purpose of the liberal arts college is preparation for life. Just what form

this preparation should take gives ample scope for many pages of argument. It would appear, however, that in one particular field at least, we could make a plea for revision of aims and methods which might secure a reasonable uniformity of opinion. More of this later.

Without doubt, Oxford in times past gave excellent preparation for life within the British Empire. One has merely to consider the great number of Oxford graduates who entered the political life of their country in one field or another to realize that this institution with its aristocratic back-ground rendered eminent service to the aristocracy which it served.

Similarly, the liberal arts college in America fifty years back undoubtedly gave excellent training along those lines for which it had been evolved. Comparatively few of its graduates entered the business and industrial life of the country The preparation and outlook which it gave for law and medicine teaching and the ministry were excellent in every respect.

Now, however, we find an entirely different state of affairs. For instance, 70 to 75 per cent of the graduates from Colgate University will, on the completion of their four-year course, immediately enter the business and industrial world. This percentage is probably high for the institutions of the country as a whole. It does however represent a very definite and emphatic shift of emphasis. The writer is Director of Placement and Vocational Guidance at Colgate University and is also intimately connected with personnel work at that institution. Every spring he sees the usual influx of high-grade business firms combing the Senior Class in search of possible executive timber.

He is also brought face to face with this situation,—a situation which is peculiar indeed when we stop to consider that a college is an institution of learning. These business firms are frankly,—one might almost say, blatantly—uninterested in the academic record of the student. Further, they have only a passing interest in the subject of the student's major. Latin, Mathematics or Sociology are all one, practically speaking, in so far as their preference of the individual student is concerned.

What these firms seek frankly and openly is that hazy entity known as personality.

They want leaders, men capable of sound and reasoned judgment on business matters, men who will give their loyalty without question or quibble to the employing firm, men who are aggressive and have a super-abundance of dynamic energy. To be sure, scholastic ability will not count against the prospective employee, neither will a major in economics—provided that first great requisite of personality be present. On the other hand the senior may have made Phi Beta Kappa in his junior year and have a straight A record in economics, only to be quietly passed by in favor of a student with a low C record and a major in geology plus a promising personality.

Now, this is all very disconcerting to the specialist in vocational guidance at an orthodox liberal arts institution. In fact, in so far as that large group of students is concerned who will enter industry, his time and efforts are liable to be largely wasted. He can, to be sure, advise this course or that course, this professor or that professor. All the time, however, he is painfully aware that his advice is of comparatively little value; for the real secret of success lies in the development of those personality traits which no professor or no course of study can develop.

This being the case, the writer would like to stress the need for increased attention to personality development within the institutions of liberal education. The population of our arts colleges has increased tremendously during the last decade, and continues to increase, although at a somewhat slower rate. We no longer serve the selected few of Oxford or of the older American institution. Our scope has broadened, our field of service has broadened; our methods of attack must be altered to meet this increasing responsibility. Our colleges must prepare for life in a democracy and suit this preparation to democratic demands.

Possibly the most important, although by no means the only fundamental modification in our line of attack must be the development of character and the correction of personality diffi-

culties. The writer does not wish to overstress this point. He certainly has no intention of claiming that the liberal arts college should become purely or even to a large extent an institution for the development of those traits of personality and character which lead to happiness in life and success in the economic sphere. He would simply point out that whereas possibly a majority of liberal arts graduates at the present time are entering business, and whereas business success depends so greatly on personality and to such a relatively small extent on scholastic eminence or on subjects studied, that the college is simply shirking its responsibilities if it does not at least attempt to furnish the student with the preparation necessary.

Guidance along these lines is essentially a matter for the expert. It demands an intimate knowledge of those personality traits required by industry and it also calls for a close acquaintance with psychology, especially with that broad field now included under the hazy concept of mental hygiene.

The writer anticipates that this particular college department, if we may call it such, will rapidly develop in the near future to be one of the most important phases of the college teaching staff. It is, if you will, vocational guidance, but with a new slant. It resembles personnel work, but is widely different from the usual idea of that term. Indeed, this particular type of service to the student body seems to the writer to differ quite markedly from any now rendered to the college population on a systematic basis. Such a service must, however, be one of the outstanding features in that modified liberal arts institution which will be the result of our present processes of change and evolution.

NEWS AND NOTES

The results of a recent survey made by the Education Division of the National Safety Council show that safety education has become generally accepted as an important subject in the elementary schools. Eighty-six per cent of the school systems replying to a questionnaire teach safety in the elementary schools while only 56 per cent teach it in the secondary schools. In 386 cities safety is taught as a separate subject and in 760 cities it is taught as an extra-curricular subject. Other interesting facts are set forth in the tabulated results of the questionnaire and show a growing interest in the importance of safety teaching. Further information may be obtained from the Education Division, National Safety Council, 1 Park Avenue, New York City.

"Individuality Through Guidance" will be the keynote of the Eleventh Annual Ohio State Educational Conference which will be held at Columbus, April 9, 10, 11, 1931. Among the more than one hundred speakers scheduled to take part in the program will be Charles Davila, Ambassador from Chili; Michael West, Principal of Teachers College, Dacca, India; and Newton D. Baker. An attendance of 6,000 is anticipated, with thirty-six sectional divisions represented. To the readers of this JOURNAL the following sections will certainly be of much interest: adult education, clinical psychology, guidance counselors, educational and intelligence tests, industrial and vocational education and special education.

The World Book Company is publishing a new test by Otis and Orleans, Standard Graduation Examination, which is to be used as a final examination in all subjects for pupils completing the elementary school. The test consists of questions covering ten different subjects: spelling, reading, vocabulary, arithmetic computation, arithmetic reasoning, grammar, language usage, literature, history and geography. A new form will be published each year and distributed just before the close of schools only on order from school administrators or authorized agents. The authors of this examination claim for it many advantages over the traditional-type final examination. It requires much less time to give and mark; it contains subject matter applicable to all parts of the coun-

try; it is entirely objective. Orders must be placed by May 1, 1931. Price per package of 25 copies, $1.25.

The Office of Education, Department of the Interior, announces the publication of a bibliography of 3,065 investigations by 317 research agencies covering the school year 1928–1929. It gives information on 217 Doctors' theses and 1,209 Masters' theses. This bibliography is a continuation of work done by Dr. Walter S. Munroe, University of Illinois, from 1917 to 1927. Copies of the bibliography may be obtained from the Superintendent of Documents at 45 cents each.

The Ohio Academy of Science will hold its forty-first annual meeting on April 2, 3, and 4 at Miami University, Oxford. A cordial invitation has been extended to both the Indiana and the Kentucky Academies to meet with the Ohio Academy at this time and place. In general the program will consist of a field trip or two on Thursday afternoon, April 2, to points of interest in the immediate vicinity of Oxford followed that evening by an informal address with discussion and a good social time. On Friday morning there will be a general session of the three Academies at which the presidents, or persons selected by them, will make 30-minute addresses following which the Academies will break up into sectional meetings. These sectional meetings will continue through the afternoon. On Friday evening the annual dinner will take place at which many interesting things always occur—speeches, music, delightful social hour, etc. Saturday will be given over largely to sectional meetings possibly with another field trip for those who wish to go. This meeting is unusual and is sure to be of great interest.

Dr. Ross O. Runnels, Director of Research for the schools of Maplewood and South Orange, N. J., has published recently a table of equivalent scores and mental ages for five widely used intelligence tests: National, Haggerty, Otis Advanced, Terman Group, and Otis Self-Administering. Comparability of norms is becoming a matter of major importance as testing programs grow. This manual, Test Method Helps, No. 2, can be secured from the World Book Company; the price, 10 cents.

According to a statistical study just completed by Albert W. Whitney, associate general manager of the National Bureau of Casualty and Surety Underwriters of New York, it has been found that the number of adult accidental fatalities has been increasing each year while that for children has steadily decreased. School teachers, policemen assigned to traffic duty near schools and those who are active in safety work are saving the lives of 6,000 children a year and preventing the injury of many thousands more. "Similar reductions in the fatality rate among adults could undoubtedly be effected if the same means of educating

older people were available," says Mr. Whitney. "This year the number of fatalities caused by accidents will probably reach 100,000, the greatest single cause of dependency and destitution."

Announcements have been received of the First International Congress of Religious Psychology to be held in Vienna, May 25–30, 1931. Those becoming members by paying the annual fee of $3.00 will receive the Quarterly "Zeitschrift fur Religions-psychologie" and all reports and announcements concerning the meetings. Those interested may communicate with Professor Karl Beth, Wien VII, Zitterhofergasse 8, Austria.

At a dinner held on January 14, arranged through the coöperation of The American Woman's Association and The National Committee for Mental Hygiene, the new General Director of the National Committee, Dr. C. M. Hincks, made his first public address. Other speakers included Dr. Marion E. Kenworthy, psychiatrist for the Child Guidance Institute; Dr. Arthur R. Ruggles, superintendent and psychiatrist of the Butler Hospital in Providence, R. I.; and Mr. Clifford Beers.

BOOK REVIEWS

CHARLES W. ST. JOHN. *Educational Achievement in Relation to Intelligence.* Harvard Studies in Education, Harvard University Press, Cambridge, Mass., 1930. 219 pp. $3.50.

The title of the present study is qualified by the author as "Educational Achievement in Relation to Intelligence as Shown by Teachers' Marks, Promotions and Scores in Standard Tests in Certain Elementary Grades." The author very carefully avoids the assumption that he is measuring the relation between intelligence and *all* the outcomes of education. The study comprises a considerable body of original experimental work with approximately a thousand school children in the lowest six grades over a period of four years. In addition there is a review of a number of previous studies carried out in the same field. The experimental work done here seems to have been undertaken with more thoroughness than most previous studies but the general results are much the same. Correlation techniques are used with the general finding that the highest correlations are observed between scores in standard tests and intelligence. Even here, though, the author is at pains to point out that there are other factors determining school success which, in their aggregate effect, must be considered fully as important as the single one of intelligence. Sex differences in securing school marks and promotions are in favor of girls, although on standardized tests, correlations show approximately the same relation between achievement and intelligence for both sexes. Some sidelines of investigation, e.g., "repeaters," offer added interest to the reader. A few case studies of exceptional children and emphasis to the writer's warnings that intelligence is not the sole factor in determining school success. The study is cautious and the author is wary of drawing any sweeping conclusions. For the "amateur" reader in this field the author has included some explanations of techniques which the more experienced reader will hasten past. Chapters are well summarized.

STUART M. STOKE,
Ohio University.

MORDECAI EZEKIEL. *Methods of Correlation Analysis.* John Wiley and Sons, Inc., New York, 1930. xiv + 427 pp.

This volume is not a book on general statistics, but rather a presentation of methods and techniques of studying the relations between vari-

ables—a presentation of analysis by means of various correlational procedures. It is not a mathematical presentation of the subject and does not present the proof of any of the theorems presented, except in the appendix. We find it primarily a book in which are outlined practical procedures for the solution of problems, and as such, it is one that could be used profitably by almost all statisticians. The style of writing is sufficiently clear that an individual, even with rather limited statistical training, say one or two courses, can follow the techniques and explanations quite readily.

One factor which does much to make the book easy to read is the extensive use of examples which not only illustrate the application of some principle, but show how the formulas are applied, step by step, in actual problems. Most of the examples are, of course, from the field of agricultural economics, but the explanations are sufficiently general that the methods of analysis may be readily carried over and applied in other fields.

Newer methods and techniques are presented. These include such techniques as Ezekiel's approximation to the curvilinear multiple correlation, the Doolittle method of solving multiple correlation and regression, graphic approximations to curvilinear multiple correlation, the recognition of and development of methods of handling joint correlation, the studies of reliability of small samples as developed by R. A. Fisher and "Student," etc. It is one of the aims of the book to present in organized form many of the later developments of statistics, otherwise available only in articles.

At the end of each chapter is a bibliography giving references to studies in both the theory and application of principles discussed in the chapter.

The book makes an excellent handbook of the statistics of covariation, and would probably serve well as a text for a course in correlation.

HAROLD A. EDGERTON,
Ohio State University.

D. B. LUCAS AND C. E. BENSON. *Psychology for Advertisers.* Harper and Brothers, New York, 1930. xv + 351 pp. $3.50.

Many textbooks of advertising psychology have been either compendiums of experimental results, or books on the mechanics of advertising with a little psychology thrown in for good measure. This text falls into neither of these classes probably because its authors bring to the task a knowledge of practical advertising and of experimental psychology. In the introduction they say, "the aim of the book is to present the theoretical, though scientifically supported, psychological aspects of advertising. They have succeeded commendably.

The applications of psychology are pointed out in five aspects of ad-

vertising: (1) understanding the customer, (2) advertising appeals, (3) constructing the advertisement, (4) choosing the medium, and (5) measuring the results of advertising.

Part one, concerned with the customer, treats in an excellently simple, yet thorough, manner the physiological basis of behavior, factors in attention and interest, and the native and habitual drives and motives. These three chapters afford the student or intelligent advertiser with a survey of the pertinent aspects of psychology.

Part two, on the psychology of appeals, explains the relation of appeals to human motives and discusses the relative value of different types of appeals. This section is particularly well illustrated with excellent half-tone reproductions of recent advertisements.

Part three discusses the layout of the advertisement. Color, type face, position, size, illustration and coupons are some of the items treated from the point of view of their psychological value. The necessity for making the appeal and layout, in all their aspects, conform to the product so that proper feeling tones are produced, is clearly indicated.

Part four, selecting a medium, compares the merits of magazines, newspapers, direct mailing, etc. The necessity of knowing the sort of public the medium serves before deciding on the appeal or layout is pointed out. The second of the two chapters in this section discusses selling and the salesman from the point of view of an advertising medium.

Part five, measurement of effectiveness, is the one most closely related to the methodology of psychology. Statistical methods and laboratory technic are fully described and explained. A separate chapter considers the statistical treatment of number of sales and inquiries in measuring the effectiveness of an advertisement or advertising campaign.

In part six individual, sex, age and geographic differences are very adequately related to the success of advertising. The last chapter is really an essay of "advertising as a comprehensive job of public education."

This rather extended description of the contents of this book is, I believe, justifiable because the text is almost unique. It is sanely written and maintains a careful balance between theory and practice. Unreservedly it is to be recommended as a text for courses in the psychology of advertising.

C. M. Louttit,
Ohio University.

A. R. Gilliland, John J. B., Morgan and S. N. Stevens. *General Psychology for Professional Students.* D. C. Heath & Co., New York, 1930. vii + 439 pp. $3.20.

This simple little textbook "has been written to fill the specific need for a special text for professional students," but to judge from the illus-

trative material it has no specific value for any of the fields ordinarily thought of as professional. Examples of the application of psychology in law, medicine or the Church are incidental if they exist at all. It is of somewhat greater value for education, while the bulk of the illustrative material applies to business, e.g., advertising, selling, personnel, etc. The reviewer feels that it does not adequately fill the need.

From the point of view of general psychology its content is pitched at so low a level that it would hardly be acceptable as a college text. The structural basis of behavior and simple innate behavior are summarily treated, with several statements which if not erroneous are very misleading. I am tempted to quote some examples. "Integrative activity takes place as a result of some stimulation of a sense organ" (p. 62). "The largest component [innate behavior] is the morphological structure of the organism" (p. 68). "The simple reflex consists of a sense organ . . . . , a sensory nerve . . . . , and a motor fiber. . . . " (!) (p. 76).

These quotations have been taken from the second and third chapters as illustrative of the loose writing; more such excerpts are to be found in later chapters. However, the treatment improves as the book progresses and reaches its culmination in the chapter on thinking. This chapter seems to the reviewer to be excellent elementary discussion, not confused with controversial speculations.

As a textbook this work could hardly be recommended for use with college students, even Freshmen, although with some corrections it might prove useful at the high school level.

C. M. Louttit,
Ohio University.

NEW BOOKS AND PAMPHLETS RECEIVED*

Books and pamphlets for review should be sent to James P. Porter, Department of Psychology, Ohio University, Athens, Ohio.

An Anthology of Modern Philosophy. DANIEL SOMMER ROBINSON. Thomas Y. Crowell Company, New York City. $4.50. 836 pp.

Department Stores: Recent Policies, Costs and Profits. BORIS EMMET. Stanford University Press, Stanford University, California. $4.00. 214 pp.

Growth and Development of the Young Child. WINIFRED RAND, MARY E. SWEENY AND E. LEE VINCENT. W. B. Saunders Company, Philadelphia, Pa. 394 pp.

Healthy-Minded Children, The. NELSON ANTRIM CRAWFORD AND KARL A. MENNINGER. Coward-McCann, 425 Fourth Avenue, New York City. $1.75. 198 pp.

Interest and Ability in Reading. ARTHUR I. GATES. Macmillan Company, New York City. $1.50. 264 pp.

Jewish Experiences in America. BRUNO LASKER. The Inquiry, 129 E. 52nd Street, New York City. 309 pp.

The Measurement of Nervous Habits in Normal Children. WILLARD C. OLSON. University of Minnesota Press, Minneapolis, Minn. $2.00. 97 pp.

New Schools for Young India. WILLIAM J. MCKEE. University of North Carolina Press, Chapel Hill, N. C. $4.50. 435 pp.

Non-Language Primary Mental Test. RUDOLF PINTNER. Bureau of Publications, Teachers College, Columbia University, New York City, 1930.

Predicting the Scholastic Success of College Students. CHARLES W. ODELL. University of Illinois, Bulletin XXVIII, No. 5, University of Illinois, Urbana, Ill., 1930. 43 pp.

Problem Tendencies in Children. WILLARD C. OLSON. University of Minnesota Press, Minneapolis, Minn. $2.00. 92 pp.

Psychology for Teachers. ELLSWORTH COLLINGS AND MILBOURNE O. WILSON. Charles Scribner's Sons, New York. 474 pp.

Social Psychology. E. T. KREUGER AND WALTER C. RECKLESS. Longmans, Green and Company, New York City, 1931. 578 pp.

---

* Mention here does not preclude further comment.

Directory of American Psychological Periodicals (*Continued*)

Psychological Abstracts—Eno Hall, Princeton, N. J. American Psychological Association. Subscription $6.00. 700 pages ann. Edited by W. S. Hunter. Monthly. Abtracts of psychological literature. Founded 1927.

Journal of General Psychology—Worcester, Mass.: Clark University Press. Subscription $7. 500–600 pages annually. Edited by Carl Murchison. Quarterly. Experimental, theoretical, clinical, and historical psychology. Founded 1927.

Journal of Social Psychology—Worcester, Mass.: Clark University Press. Subscription $7.00. 500–600 pages annually. Edited by John Dewey and Carl Murchison. Quarterly. Political, racial, and differential psychology. Founded 1929.

Journal of Personnel Research—Baltimore: The Williams & Wilkins Co. Subscription $5.00. 500 pages approximately. Edited by Walter V. Bingham, and Clarence S. Yoakem. Bi-Monthly.

Mental Measurement Monographs—Baltimore: The Williams & Wilkins Company. Subscription $5. 400 pages annually. Edited by Knight Dunlap and Buford Johnson. Published without fixed dates, each number a single research. Founded 1925.

Studies in Psychology and Psychiatry—Baltimore: The Williams & Wilkins Company. Subscription $5. 400 pages annually. Edited by Edward A. Pace. Published without fixed dates, each number a single research. Founded 1926.

Directory of American Psychological Periodicals

American Journal of Psychology—Ithaca, N. Y.; Cornell University. Subscription $6.50. 624 pages annually. Edited by M. F. Washburn, K. M. Dallenbach, M. Bentley, E. G. Boring. Quarterly. General and experimental psychology. Founded 1887.

Journal of Genetic Psychology and The Pedagogical Seminary—Worcester, Mass.: Clark University Press. Subscription $7. 600 pages annually. Edited by Carl Murchison. Quarterly. Child behavior, differential and genetic psychology. Founded 1891.

Psychological Review—Princeton, N. J.: Psychological Review Company. Subscription $5.50. 540 pages annually. Bi-monthly. General Psychology. Edited by Howard C. Warren. Founded 1894.

Psychological Monographs—Princeton, N. J.: Psychological Review Company. Subscription $6.00 per vol. 500 pp. Ed. by Raymond Dodge. Published without fixed dates, each issue one or more researches. Founded 1895.

Psychological Index—Princeton, N. J.: Psychological Review Company. Subscription $4.00. 300-400 pp. Edited by W. S. Hunter. An annual bibliography of psychological literature. Founded 1895.

Psychological Bulletin—Princeton, N. J.: Psychological Review Company. Subscription $6.00. 720 pages annually. Psychological literature. Monthly. Edited by Edward S. Robinson. Founded 1904.

Archives of Psychology—New York: Columbia University. Subscription $6. 500 pp. ann. Ed. by R. S. Woodworth. Published without fixed dates, each number a single experimental study. Founded 1906.

Journal of Abnormal and Social Psychology—Eno Hall, Princeton, N. J. American Psychological Association. Subscription $5. Foreign, $5.25. Editor, Henry T. Moore. Quarterly. 448 pages annually. Abnormal and Social. Founded 1906.

Psychological Clinic—Philadelphia: Psychological Clinic Press. Subscription $3.00. 288 pages. Ed. by Lightner Witmer. Without fixed dates (9 Numbers). Orthogenics, psychology, hygiene. Founded 1907.

Psychoanalytic Review—Washington, D. C.: 3617 10th St., N. W. Subscription $6. 500 pages annually. Psychoanalysis. Quarterly. Ed. by W. A. White and S. E. Jelliffe. Founded 1913.

Journal of Experimental Psychology—Princeton, N. J. Psychological Review Company. 500 pages annually. Experimental. Subscription $6.00. Bi-monthly. Edited by S. W. Fernberger. Founded 1916.

Journal of Applied Psychology—Baltimore: The Williams & Wilkins Co. Subscription $5.00. 400 pages annually. Bi-monthly. Edited by James P. Porter, Ohio University, Athens, Ohio. Founded 1917.

Journal of Comparative Psychology—Baltimore: The Williams & Wilkins Co. Subscription $5.00 per volume of 450 pages. Three volumes every two years. Edited by Knight Dunlap and Robert M. Yerkes. Founded 1921.

Comparative Psychology Monographs—Baltimore: Johns Hopkins University Press. Subscription $5. 400 pages per volume. Knight Dunlap, Managing Editor. Published without fixed dates, each number a single research.

Genetic Psychology Monographs—Worcester, Mass.: Clark University Press. Subscription $7.00 per volume. Two volumes per year, 600 pages each. Edited by Carl Murchison. Monthly. Each number one complete research. Child behavior, animal behavior and comparative psychology. Founded 1925.

(Continued on Third Cover)